An Improved Lower Bound for Depth Four Arithmetic Circuits

A THESIS

SUBMITTED FOR THE DEGREE OF

Master of Science (Engineering)

IN THE

Faculty of Engineering

BY

Abhijat Sharma

Computer Science and Automation
Indian Institute of Science
Bangalore – 560 012 (INDIA)

July, 2017

Declaration of Originality

I, **Abhijat Sharma**, with SR No. **04-04-00-10-21-14-1-11590** hereby declare that the material presented in the thesis titled

An Improved Lower Bound for Depth Four Arithmetic Circuits

represents original work carried out by me in the **Department of Computer Science and Automation** at **Indian Institute of Science** during the years **2014-2017**.
With my signature, I certify that:

- I have not manipulated any of the data or results.
- I have not committed any plagiarism of intellectual property. I have clearly indicated and referenced the contributions of others.
- I have explicitly acknowledged all collaborative research and discusions.
- I have understood that any false claim will result in severe disciplinary action.
- I have understood that the work may be screened for any form of academic misconduct.

Date: Student Signature

In my capacity as supervisor of the above-mentioned work, I certify that the above statements are true to the best of my knowledge, and I have carried out due diligence to ensure the originality of the report.

Advisor Name: Advisor Signature

DEDICATED TO

My beloved family

who taught me to endure and always follow my heart.

Acknowledgements

I'd like to thank my advisor Chandan Saha for his continuous help and support during my entire stay at IISc. His precious advice and suggestions have shaped my thesis work significantly. He has been instrumental in providing direction to my research, and getting me back on track in times when I got distracted. My colleagues at the Complexity Theory Lab have always been an immense help. The countless group-discussions and presentations provided valuable contribution to my work. Especially, I can never thank Vineet enough for showing great patience and consideration while revising my final submission. Sumant has been my neighbour in the lab, and I have always drawn inspiration and motivation from seeing him work hard.

I have been very fortunate to have a large group of friends at IISc, some at CSA and some at my hostel, who made my stay memorable and supported me through the thick and thin. My fellow 'theoreticians' Mayank, Datta, Giri and Indranil, and other CSA mates Akhil, Saravanan, Aritra, Shyam and Sneha, have all been great company as we strived towards our respective life goals. I must also mention the valuable philosophical discussions with Rupam, Abhishek and Gulshan, carrying me through sleepless nights in the hostel. Bangalore is a lovely city and I spent some good times here with my friends from IIIT-H, Utsav, Mudit and Kartik, who were there whenever I needed a little break from work.

I would like to thank my professor from IIIT-H, Kannan Srinathan, who was my first inspiration to pursue research in mathematics and theoretical computer science, and has been a guiding light ever since. Finally and most importantly, I convey my regards and gratitude to all the members of my family who have been my supreme source of confidence. I was always inspired to pursue a career in academics, by my Dadaji and Nanaji, who have been elite professors of science. The vast amount of faith and blessings that I receive from Dadi, Nani and all my elders, always lifts my spirits up and keeps me working hard to make them proud. My salutations to my parents, my pillars of support who always believed in me, and my little sister Anu for those chirpy little phone calls that worked as instant stress-relief.

Abstract

We study the problem of proving lower bounds for depth four arithmetic circuits. Depth four circuits have been receiving much attraction when it comes to recent circuit lower bound results, as a result of the series of results culminating in the fact that strong enough lower bounds for depth four circuits will imply super-polynomial lower bounds for general arithmetic circuits, and hence solve one of the most central open problems in algebraic complexity i.e a separation between the VP and VNP classes. However despite several efforts, even for general arithmetic circuits, the best known lower bound is $\Omega(N \log N)$ by [BS83], where N is the number of input variables. In the case of arithmetic formulas, [Kal85] proved a lower bound that is quadratic in the number of input variables, which has not seen any improvement since then. The situation for depth three arithmetic circuits was also similar for many years, until a recent result by [KST16] achieved an almost cubic lower bound that improved over the previous best quadratic bound by [SW99].

As the main contribution of this thesis, we prove an $\tilde{\Omega}(N^{1.5})$ lower bound on the size of a depth four circuit, for an explicit multilinear N-variate polynomial in VNP with degree $d = \tilde{\Theta}(\sqrt{N})$. Our approach offers a potential route to proving a super-quadratic lower bound for depth four circuits. Taking cue from the numerous successful results recently, we use the technique of the shifted partial derivatives measure to achieve the said lower bound. Particularly, we use the Dimension of Projected Shifted Partials (DPSP) measure which has been previously used in [KLSS14] and [KS15]. Coming to the choice of the hard polynomial, we again follow the status quo and use a variant of the Nisan-Wigderson (NW) polynomial family that has proved to be very helpful over the past few years in arithmetic circuit complexity.

Finally, we do a careful analysis of [SS97] and [Raz10] and compare their techniques to ours. We conclude that our result can potentially be used as a starting point, and techniques similar to [KST16] can likely be used to strengthen our lower bound to $\tilde{\Omega}(N^{2.5})$, for general depth four arithmetic circuits. However, unlike depth three circuits, proving a super-quadratic lower

bound for depth four circuits remains a prevalent open problem for many years. Previous work like [SS97] and [Raz10] implied super-linear lower bounds. To the best of our knowledge, the previous best known lower bound for general depth four circuits is $\tilde{\Omega}(N^{1.33})$.

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

Proving lower bounds for various interesting computational models is one of the most important aspect of computational complexity theory. The famous P versus NP problem (introduced formally in [Coo71]) can be seen as an instance of a lower bound proving problem. Viewing computation on Turing machines as Boolean circuits, it becomes sufficient to prove super-polynomial lower bound for boolean circuits computing an NP-function, in order to show $\mathsf{P} \neq \mathsf{NP}$. With boolean circuits came the notions of *circuit-size* and *circuit-depth* complexity, and the corresponding complexity classes $\mathsf{AC}^0, \mathsf{AC}, \mathsf{NC}$ etc.

Arithmetic circuits ([Val79],[Val82]) are arithmetic analogues of boolean circuits, and the area which studies this model is known as arithmetic circuit complexity. Like Boolean circuits, arithmetic circuits form a non-uniform model of computation and they are structurally very similar to boolean circuits. Arithmetic circuits are acyclic directed graphs just like their boolean counterparts, only the OR and AND gates are replaced by *sum* $(+)$ and *product* $(\times)$ gates respectively. The leaves are still labelled by one of the input variables $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ or a constant from the underlying field $\mathbb{F}$. The edges are also sometimes labelled by field constants so the $+$ gate eventually computes a weighted sum of its children, and similarly the $\times$ gate computes the weighted product of its children. Thus it can be realised that an arithmetic circuit basically captures the step-by-step computation of a multivariate polynomial in $\mathbb{F}[\mathbf{x}]$, and so it is also called a *straight-line program*. Similar to boolean circuits, there are two measures associated with an arithmetic circuit: the *size* which is defined as the total number of edges, and the *depth* which is the length of the longest directed path in the circuit. The formal description of an arithmetic circuit and associated parameters are discussed in Section 2.1.

Valiant ([Val79], [Val82]) defined two complexity classes (for the arithmetic world) analogous to the P and NP complexity classes, called as VP and VNP. VP is defined as the class of polynomial families that can be computed by polynomial-sized arithmetic circuits. VNP can be defined analogous to the "polynomial-time verifiable" definition of NP, as the class of polynomial families that given any monomial can determine its coefficient in polynomial time. As VP and VNP were defined in close similarity to their analogs P and NP, naturally it gave rise to the corresponding important open question in algebraic complexity, is $\mathsf{VP} \neq \mathsf{VNP}$? Just like $\mathsf{P} \subseteq \mathsf{NP}$, the containment $\mathsf{VP} \subseteq \mathsf{VNP}$ holds here, and the question is only whether it is a proper containment. These complexity classes are also more precisely described in Section 2.1. Hereafter in the thesis, by circuit(s) we shall mean arithmetic circuit(s).

1.2 Constant Depth Circuits and Motivation

The depth of a circuit corresponds to the amount of parallel time spent to compute a polynomial using the circuit. So, lower the depth of circuits computing a polynomial, the faster it is to compute the polynomial in parallel. In a breakthrough result on depth reduction by Valiant et. al. ([VSBR83]) proved that if a polynomial f having total degree d can be computed by an arithmetic circuit of size s, then f can also be computed by a circuit of depth $O(\log d)$ with size only polynomial in s and d. This was followed by a series of improvements ([AV08],[Koi12], [Tav13]) that utilized the $O(\log d)$ depth circuit to construct a depth four circuit computing the same polynomial f while keeping the size still sub-exponential. The structure of the reduced arithmetic circuit is such that the sum and the product gates are arranged in alternating layers as follows: the output gate is a sum gate, with all its children being product gates and so on. Therefore, these depth four circuits are commonly represented as $\Sigma\Pi\Sigma\Pi$ circuits, Σ and Π denoting layers of sum gates and product gates respectively. The depth four circuits resulting from depth reduction are also homogeneous, meaning all gates in the circuit compute homogeneous polynomials. Eventually, the series of research ([AV08],[Koi12], [Tav13]) concluded that any N-variate d-degree polynomial $f \in \mathsf{VP}$ can be computed by a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit of size $N^{O(\sqrt{d})}$ and bottom fan-in $O(\sqrt{d})$. Here, the bottom fan-in refers to the maximum fan-in of the gates in the layer closest to the input variables. Thus, it can be concluded that a $N^{\omega(\sqrt{d})}$ lower bound on the size of any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit (in-fact a low-bottom fan-in homogeneous depth four circuit) computing a polynomial in VNP, would imply that $\mathsf{VP} \neq \mathsf{VNP}$. Interestingly, this direction has seen significant progress over the recent years through the results of [Kay12], [GKKS13b], [KSS14], [FLMS14], [KLSS14] and [KS14] which finally proved a $N^{\Omega(\sqrt{d})}$ lower bound for depth four homogeneous circuits (without any bottom fan-in restric-

tion), computing the iterated matrix multiplication (IMM) polynomial.

In a surprising result, the $\Sigma\Pi\Sigma\Pi$ circuit from [AV08] was further reduced to a depth three ($\Sigma\Pi\Sigma$) circuit in [GKKS13a]. Unlike depth four, the reduction to depth three circuit does not preserve homogeneity, but still has bottom fan-in bounded by $O(\sqrt{d})$. Like depth four, it can be concluded that an $N^{\omega(\sqrt{d})}$ lower bound for depth three circuits with low bottom fan-in implies general circuit lower bound. Building on the work of [KLSS14], [KS15] showed an $N^{\Omega(sqrtd)}$ lower bound for depth three circuits with low bottom fan-in computing a polynomial in VNP (which was improved to a polynomial in VP, namely IMM, in [BC15]).

The lower bounds for homogeneous depth four and of depth three circuit with low bottom fan-in have reached a threshold crossing which would imply $\mathsf{VP} \neq \mathsf{VNP}$. It is natural to ask if the current techniques can be used to at first prove super-polynomial lower bounds for general depth three and depth four circuits. Proving super-polynomial lower bounds for depth three circuits has been posed as an open problem (among many other places) in [Wig06], in terms of the determinant polynomial as Open Problem 7.5. For quite some time the best known lower bound at depth three (for circuits as well as formulas) was quadratic ([SW99]). Attempting to prove super-quadratic lower bound even at depth three or four is interesting as the best known lower bound for general formulas is quadratic ([Kal85]). So, these low depth models (with no other restriction like homogeneity or low bottom fan-in) serve as interesting testbed to strengthen our techniques to achieve super-quadratic bounds. Recently, there has been some progress for general depth three circuits/formulas in the form of a modest (almost) cubic lower bound ([KST16], [BLS16], [Yau16]). Can we achieve a similar super-quadratic lower bound for general depth four circuits/formulas? This has also been mentioned in [SY10], (Section 1.4.2, page 8). This question serves as the main motivation behind our work.

Most of the recent depth four lower bound results are based on the concept of *partial derivative measures*. Nisan and Wigderson ([NW97]) introduced the use of partial derivatives in the arithmetic circuit lower bound literature, to prove an exponential lower bound for homogeneous depth three circuits. Kayal ([Kay12]) then came up with an augmentation of the partial derivatives measure, called *the shifted partials dimension*, followed by by multiple extensions and alterations of the measure to achieve more lower bound results. In our proof, we employ one such variation of the measure, known as *dimension of the projected shifted partial derivatives* (DPSP) that was also first used in [KLSS14]. Partial derivatives based techniques have been an essential part of analyzing arithmetic circuits, and the interested reader can explore

the extensive use of partial derivatives in the survey [CKW11].

1.3 Previous Work

As mentioned before, we have seen a lot of progress in lower bounds for homogeneous depth four circuits, multilinear depth four circuits, depth four circuits with restricted bottom fan-in etc. But when it comes to general $\Sigma\Pi\Sigma\Pi$ circuits without any restrictions, there has not been much progress in recent times. We record some relevant results in this regard. Baur and Strassen ([Str73],[BS83]) obtained a $\Omega(N \log d)$ lower bound on the size of a circuit of any depth, where d is the degree of the polynomial being computed. There are known super-linear bounds for constant-degree multilinear polynomials (for instance, a polynomial derived from a product of two symbolic matrices). First such bounds were proved by [Pud94] and later [RS03], which are asymptotically super-linear. For depth-Δ circuits, these bounds are of the form $\Omega(N.\lambda_\Delta(N))$, where $\lambda_\Delta(N)$ are extremely slowly growing functions ($\lambda_\Delta(N) \ll \log N$).

An improvement over this was the polynomial interpolation lower bound by Shoup and Smolensky ([SS97]). They proved a lower bound of $\Omega(\Delta.N^{1+1/\Delta})$ for depth-Δ circuits computing an explicit polynomial of degree $O(N)$. Later, Raz proved an $N^{1+\Omega(1/\Delta)}$ lower bound on the size of any depth-Δ arithmetic circuit computing an explicit polynomial of degree $O(\Delta)$ in $O(n\Delta)$ variables. The motivation for considering explicit polynomials of constant degree comes from the fact that super-quadratic lower bound for constant depth circuits computing an explicit polynomial of constant degree will imply super-linear lower bound for general (depth unrestricted) circuits [Raz10].

The above mentioned results [SS97] and [Raz10] imply super-linear lower bounds for depth four circuits. We discuss both the above works in more detail, and analyse their techniques as compared to our $\tilde{\Omega}(N^{1.5})$ lower bound for depth four circuits computing an explicit $\tilde{\Theta}(\sqrt{N})$ degree polynomial, in Chapter 5. However, the question of proving a super-quadratic lower bound for depth four circuits (over fields of characteristic other than two) remains open (mentioned in the survey [SY10]). Nevertheless, as stated before, our approach does have the promise of leading to a super-quadratic lower bound.

1.4 Our Contribution

The structure and computation process of a general $\Sigma\Pi\Sigma\Pi$ circuit C can be formally described as follows:

$$C = T_1 + \ldots + T_s \text{ and } T_i = \prod_{j=1}^{s_1} P_{ij} \tag{1.1}$$

where s is the top fan-in of the $\Sigma\Pi\Sigma\Pi$ circuit, every T_i ($i \in [s]$) is a $\Pi\Sigma\Pi$ circuit (s_1 being the maximum of the top fan-ins of all the $T_1, \ldots, T_s$), and P_{ij}'s are polynomials with sparsity (total number of monomials) bounded by s_2 (say). In the rest of this thesis, we shall assume that the underlying field has characteristic zero. The main lower bound result for $\Sigma\Pi\Sigma\Pi$ circuits proved in this work, is as follows.

Theorem 1 *There exists an explicit family of N-variate degree d polynomials, $\{f_N\} \in \mathsf{VNP}$, such that any $\Sigma\Pi\Sigma\Pi$ circuit, over any field of characteristic zero, computing f_N must have size $\Omega(\frac{Nd}{\log^5 N})$ for any $d \leq \frac{\sqrt{N}}{\log^4 N}$.*

In the process of proving Theorem 1, we first enforce the constraint,

$$s_1, s_2 \leq \left(\frac{Nd}{\log^5 N}\right),$$

otherwise the target lower bound as in Theorem 1 would be proved already. Hence, the theorem can be restated more precisely as the following lemma.

Lemma 1.1 *There exists a family of N-variate, degree d polynomials $\{f_N\}$ in VNP such that for any $\Sigma\Pi\Sigma\Pi$ circuit C that computes f, if $s_1, s_2 \leq (\frac{Nd}{\log^5 N})$ (where s_1, s_2 are the fan-in bounds on the level 3 product gates and level 2 sum gates respectively), then the top fan-in of C, $s = \Omega(\frac{Nd}{\log^5 N})$.*

We give more explicit details about the hard polynomial family $\{f_N\}$ in later chapters, but here we substitute the degree of the polynomial f_N, $d = \frac{\sqrt{N}}{\log^4 N}$ to achieve our main result as a corollary of Theorem 1.

Corollary 1.2 *There exists a family of N-variate, degree $d = \frac{\sqrt{N}}{\log^4 N}$ polynomials $\{f_N\}$ in VNP such that any $\Sigma\Pi\Sigma\Pi$ circuit, over any field of characteristic zero, computing f_N must have size $\tilde{\Omega}(N^{1.5}) = \Omega(\frac{N^{1.5}}{\mathsf{poly}(\log N)})$.*

We will proceed to prove Lemma 1.1 in two steps (closely following the recent literature),

- proving an upper bound on the measure *Dimension of Projected Shifted Partials* ($\mathsf{DPSP}_{k,\ell}$) for a general depth four circuit C (under a random restriction), for a certain choice of the parameters k, ℓ,

- proving a lower bound on the same measure used above for the hard polynomial family $\{f_N\}$, for the same parameters k, ℓ.

We use random restrictions on the circuit C to restrict the degree of all monomials in C, by a fixed bound. This makes the restricted circuit simpler and easier to analyse, and obtain the required upper bound on the measure.
The explicit polynomial is an instance of the Nisan-Wigderson (NW) polynomial family. The lower bound is obtained by closely following the arguments in [KLSS14] and [KS15]. We define a matrix M, such that the rank of M is exactly equal to the required measure. Then, the problem of lower bounding rank(M) is simplified by constructing another matrix B such that rank$(M) \geq Rank(B)$ and rank(B) is easier to lower bound through the concept of *Surrogate Rank* ([Alo09], and just as in [KLSS14]), which we discuss in detail later.

1.5 Organisation

We start by introducing some basic notations and definitions in Chapter 2, necessary to get acquainted with the commonly used terminology related to arithmetic circuits, formulas, polynomials being computed, the complexity measure etc. In chapter 3, we prove the upper bound on the complexity measure $\mathsf{DPSP}_{k,\ell}$ for a general depth four circuit C. Chapter 4 introduces the explicit hard polynomial f, followed by proof of the lower bound $\mathsf{DPSP}_{k,\ell}(f)$ and ultimately completing the proof of Theorem 1. We also discuss possible future directions and improvement of our result. Finally, in Chapter 5, we discuss two previous lower bound results for general depth four circuits and how our result stands in their perspective.

Chapter 2

Preliminaries

In this chapter, we introduce some preliminary notations and definitions that we would be using hereafter, in the subsequent chapters.

2.1 Basic Definitions and Notations

Throughout the article, N and d denote the total *number of variables* and the *degree* of the polynomial (or the circuit depending upon context) respectively. Some other notations used in the text are:

- The set $\{1, 2, 3, \ldots, a\}$ is denoted simply as $[a]$, for any $a \in \mathbb{N}$.
- The combinatorial notation $\binom{[a]}{b}$ refers to the set of all subsets of $[a]$ of size b, for any $a, b \in \mathbb{N}$.

The main model of computation and related terms are formally defined below.

Definition 2.1 (Arithmetic Circuit) *An arithmetic circuit, defined over a set of input variables $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ and an underlying field $\mathbb{F}$, is a directed acyclic graph where each vertex (called a gate) is labelled by either an input variable, a field element, or by one of the two symbols $+$ (sum gate) and $\times$ (product gate). The incoming edges into a sum gate are also labelled with field elements, so that it computes a linear combination of its inputs. The vertex with no outgoing edges is known as the output gate as that computes the final output of the circuit, the polynomial $f \in \mathbb{F}[\mathbf{x}]$.*

An arithmetic circuit in which every gate has at most one outgoing edge is called a *arithmetic formula*. For every circuit or a formula, two parameters are defined, namely the *size* and the *depth*. The *size* of an arithmetic circuit (or formula) is defined as the number of edges in the

circuit. The *depth* of a circuit (or formula) is defined as the length of the longest path from a leaf node (labelled by an input variable or a field element) to the output gate. For our calculations, we will be primarily concerned with depth four arithmetic circuits.

As discussed briefly in Chapter 1, Valiant defined two classes in arithmetic circuit complexity:

Definition 2.2 (Class VP) *A family of multivariate polynomials $\{f_N\}$, is said to be in* VP*, if there exists some polynomial $t : \mathbb{N} \mapsto \mathbb{N}$ and an arithmetic circuit C computing f_N such that for every N, the number of variables in f_N, the degree of f_N and the size of C, are all bounded by $t(N)$. These are also called p-computable families. Thus,* VP *consists of all p-computable families of polynomials.*

Definition 2.3 (Class VNP) *A family of multivariate polynomials $\{f_N\}$, is said to be in* VNP*, if there exist two polynomially bounded functions $t, k : \mathbb{N} \mapsto \mathbb{N}$ and a family $\{g_N\} \in$ VP, such that for every N,*

$$f_N(x_1, x_2, \ldots, x_{k(N)}) = \sum_{w \in \{0,1\}^{t(N)}} g_{t(N)}(x_1, \ldots, x_{k(N)}, w_1, \ldots, w_{t(N)}) \tag{2.1}$$

These are also called p-definable families. Thus, VNP *consists of all p-definable families of polynomials.*

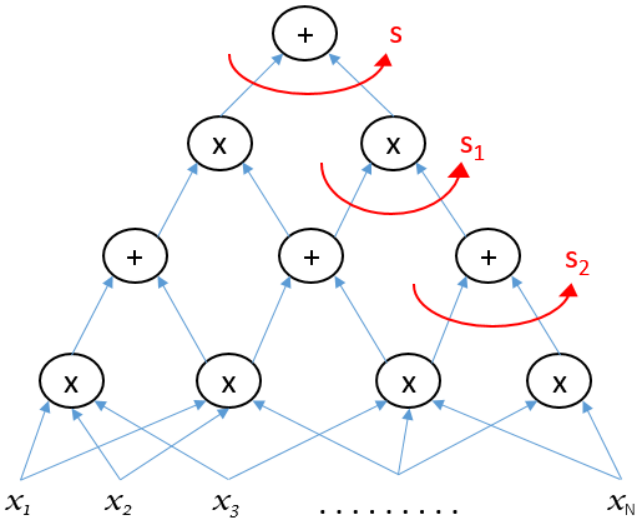

Figure 2.1: Depth four ΣΠΣΠ Circuit

Constant Depth Circuits: In this article, we work with arithmetic circuits having constant depth i.e independent of N and d. In particular, our result is a lower bound on the size of a depth four circuit, the model as depicted in Figure 2.1 above. We shall also use the term *levels* to refer to all nodes at a particular depth. The leaf nodes would be taken as level zero, the monomials being computed at level 1 product gates, and similarly the output gate would be at level 4.

In addition to the above, we shall also use the term *support* in two different contexts, defined as follows:

Definition 2.4 (Support of a monomial) *For any monomial $m \in \mathbb{F}[\mathbf{x}]$ we define the* **support** *of m, denoted as* $\mathrm{Supp}(m)$*, as the minimal set of variables that are required to be non-zero for m to be non-zero.*

$$\mathrm{Supp}(m) := \{x_i \ : \ m\mid_{x_i=0} = 0\}$$

In case of multilinear monomials, if $S = \mathrm{Supp}(m)$, then the monomial m is sometimes denoted using its support set as $\mathbf{x}_S$.

Definition 2.5 (Support of a polynomial) *The support of a polynomial f, denoted simply as* $\mathrm{Support}(f)$*, is just the number of monomials in the polynomial with non-zero coefficients.*

2.2 Random Restrictions

Given a set $S \subseteq [N]$ that refers to a subset of variables from $\{x_1, x_2, \ldots, x_N\}$, we define a substitution map $\sigma_S : \mathbf{x} \mapsto \mathbb{F} \cup \mathbf{x}$ such that:

$$\sigma_S(x_i) := \begin{cases} 0 & \text{if } i \in S, \\ x_i & \text{otherwise.} \end{cases}$$

This substitution map can be extended naturally to polynomials and sets of polynomials. Thus, for any polynomial $f \in \mathbb{F}[\mathbf{x}]$ and any set of polynomials $A \subseteq \mathbb{F}[\mathbf{x}]$,

$$\sigma_S(f) := f\mid_{x_i=0 \ \forall \ i\in S},$$
$$\sigma_S(A) := \{f\mid_{x_i=0 \ \forall \ i\in S} : f \in A\}.$$

In our proof later, we will choose the set S randomly and use the restriction σ_S to eliminate higher degree terms from the circuit. This operation is called as *random restriction*.

Definition 2.6 (Random Restriction) *A random restriction σ_R is a substitution map corresponding to a subset of variables $R \subseteq [N]$, which is obtained by picking each variable independently at random with probability $(1-p)$ and setting it to zero (every variable survives with probability p). The set R consists of all the variables that are set to zero.*

2.3 Complexity Measure: Projected Shifted Partials

We use the following complexity measure in our proof: *Dimension of the Projected Shifted Partial Derivatives*. We now proceed to describe the measure taking one term at a time.

Dimension: We shall use the notation $\dim(A)$, for any set of polynomials $A \subseteq \mathbb{F}[\mathbf{x}]$, to denote the linear dimension of the $\mathbb{F}$-linear span of the polynomials in A. It might also sometimes be referred simply as the $\mathbb{F}$-linear dimension of A.

Projection: Define a map $\pi : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}[\mathbf{x}]$ such that for any polynomial f, $\pi(f)$ retains only and exactly the multilinear monomials of f. The definition can be extended to any set of polynomials $A \subseteq \mathbb{F}[\mathbf{x}]$,

$$\pi(A) := \{\pi(f) \ : \ f \in A\}.$$

Observation 2.7 *For any set of polynomials $A \subseteq \mathbb{F}[\mathbf{x}]$, $\dim(\pi(A)) \leq \dim(A)$.*

Proof: Let $p = \dim(A)$ and $a_1, a_2, \ldots, a_p$ be a basis of A ($a_i \in A$). Then we can observe that the polynomials $\pi(a_1), \pi(a_2), \ldots, \pi(a_p)$ form a spanning set for $\pi(A)$. The above can also be inferred from the fact that $\pi(a+b) = \pi(a) + \pi(b)$ for any two polynomialss a and b. Hence, $\dim(\pi(A)) \leq p$. □

Shift: Let $\mathbf{x}^{=\ell}$ denote the set of all multilinear monomials in $x_1, \ldots, x_N$ of degree exactly ℓ. When this set is multiplied to any fixed polynomial f of degree d, we get a set of polynomials $\mathbf{x}^{=\ell}.f$, all of degree $d+\ell$ and we say that f is *shifted* by ℓ.

Partial Derivatives: Now, we define the notion of partial derivative of a polynomial f with respect to a monomial m, which is also referred to as a *k-th order partial derivative* where k is the degree of the monomial m. First we define *first order partial derivative* of f, that is $\frac{\partial f}{\partial x_i}$

where x_i is an input variable. It is defined as follows:

$$\frac{\partial f}{\partial x_i} = \sum_{m \in \text{Support}(f)} \frac{\partial m}{\partial x_i}, \tag{2.2}$$

where

$$\frac{\partial m}{\partial x_i} = \begin{cases} \frac{m}{x_i} & \text{if } m = x_i.g \text{ for some polynomial } g \in \mathbb{F}[\mathbf{x}], \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

The k-th order partial derivatives are obtained by applying the above first order derivative, to an already computed $(k-1)$-th order partial derivative, as follows:

$$\frac{\partial f}{\partial m} = \frac{\partial f}{\partial (x_i.m')} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial m'}\right),$$

where m and m' are monomials of degrees k and $k-1$ respectively. In our work, we consider derivatives with respect to multilinear monomials. We shall use $\partial^{=k} f$ to refer to the set of all multilinear k-th order partial derivatives of $f \in \mathbb{F}[\mathbf{x}]$.

Also, $\mathbf{x}^{=\ell}.\partial^{=k} f$ denotes the set of all polynomials of the form $m.g$ where $m \in \mathbf{x}^{=\ell}$ and $g \in \partial^{=k} f$. Then, for any polynomial f, the measure $\mathsf{DPSP}_{k,\ell}$ is defined as:

$$\mathsf{DPSP}_{k,\ell}(f) := \dim(\pi(\mathbf{x}^{=\ell}.\partial^{=k} f)) \tag{2.4}$$

Lemma 2.8 (Sub-additivity) DPSP *is a sub-additive measure, i.e for any fixed* k, ℓ *and any two polynomials* f, g*:*

$$\mathsf{DPSP}_{k,\ell}(f+g) \leq \mathsf{DPSP}_{k,\ell}(f) + \mathsf{DPSP}_{k,\ell}(g)$$

2.4 Nisan-Wigderson polynomials

To achieve the required lower bound, we define a polynomial f, such that $\mathsf{DPSP}_{k,\ell}(f)$ is as high as possible. One possible candidate is the Nisan-Wigderson (NW) polynomial, first introduced in [KSS14] and has been useful in proving lower bounds for certain constant depth, in particular depth three and depth four, circuits in the past few years. It is a degree-d set-multilinear polynomial on $N = d.q$ variables, where q is power of a prime. The explicit form of the NW

polynomial is given below, where r is an associated parameter.

$$\mathrm{NW}_r(x_{1,1}, x_{1,2}, \ldots, x_{d,q}) := \sum_{\substack{h(z)\in\mathbb{F}_q[z] \\ deg(h)\leq r}} \prod_{i\in[d]} x_{i,h(i)}. \tag{2.5}$$

Observe that when $q = d^\alpha$, $N = d^{\alpha+1}$ where α is any constant. We will be using an appropriate version of the above NW polynomial in proving our depth four lower bound result. To get an estimate of the 'hardness' of NW polynomial, we state the following result from [KLSS14] as a fact (without proof) below.

Fact 2.9 *Let $q = d^2, r = d/3, k = o(d), \ell = \frac{N}{2}.(1 - k\frac{\log d}{d})$. Then,*

$$\mathsf{DPSP}_{k,\ell}(\mathrm{NW}_r) \geq \frac{1}{d^{O(1)}}. \min\left(\binom{N}{\ell+d-k}, \binom{d}{k}^2.d^k.k!.\binom{N}{\ell}\right). \tag{2.6}$$

2.5 Approximations and Numerical Estimates

Stirling's Formula Stirling's formula is the following approximation:

$$\ln(n!) = n\ln n - n + O(\ln n) \tag{2.7}$$

The above is used to obtain the following estimates (proved in Section 2.6)

Lemma 2.10 *Let $a(n), f(n), g(n) : \mathbb{Z}_{>0} \mapsto \mathbb{Z}$ be integer valued functions such that $f+g = o(a)$. Then,*

$$\ln\frac{(a+f)!}{(a-g)!} = (f+g)\ln a \pm O\left(\frac{f^2+g^2}{a}\right). \tag{2.8}$$

Combinatorial Approximation The following is an approximation of the binomial coefficient that we can use either as an upper or a lower bound depending on the requirements. For any $m \geq n \geq 0$,

$$\left(\frac{m}{n}\right)^n \leq \binom{m}{n} \leq \left(\frac{me}{n}\right)^n. \tag{2.9}$$

2.6 Proof of Preliminaries

2.6.1 Proof of Lemma 2.10

Proof:

$$\frac{(a+f)!}{(a-g)!} = (a+f)(a+f-1)\dots(a-g+1) \tag{2.10}$$

$$\implies a^{f+g}\left(1-\frac{g}{a}\right)^{f+g} \leq \frac{(a+f)!}{(a-g)!} \leq a^{f+g}\left(1+\frac{f}{a}\right)^{f+g} \tag{2.11}$$

Taking logarithms on all sides,

$$\implies (f+g)\ln\left(1-\frac{g}{a}\right) \leq \ln\frac{(a+f)!}{(a-g)!} - (f+g)\ln a \leq (f+g)\ln\left(1+\frac{f}{a}\right). \tag{2.12}$$

Using the fact that $\frac{x}{1+x} \leq \ln(1+x) \leq x$ for $x > -1$, we can bound both L.H.S and R.H.S by $O\left(\frac{f^2+g^2}{a}\right)$. □

2.6.2 Proof of Lemma 2.8

Proof: Recall from the definition of DPSP:

$$\mathsf{DPSP}_{k,\ell}(f+g) := \dim(\pi(\mathbf{x}^{=\ell}.\partial^{=k}(f+g))) \tag{2.13}$$

For any monomial m of degree k, $\partial(f+g)/\partial m = \partial f/\partial m + \partial g/\partial m$. Further, for any 2 polynomials f and g, $\pi(f+g) = \pi(f) + \pi(g)$. The above two properties reduce Equation 2.13 to:

$$\mathsf{DPSP}_{k,\ell}(f+g) = \dim(\pi(\mathbf{x}^{=\ell}.\partial^{=k}(f)) + \pi(\mathbf{x}^{=\ell}.\partial^{=k}(g))) \tag{2.14}$$

This implies that every polynomial in the set $\pi(\mathbf{x}^{=\ell}.\partial^{=k}(f+g))$ is a sum of two polynomials, one each from the sets $\pi(\mathbf{x}^{=\ell}.\partial^{=k}(f))$ and $\pi(\mathbf{x}^{=\ell}.\partial^{=k}(g))$ respectively. We know that the former is generated by a basis of size $\mathsf{DPSP}_{k,\ell}(f)$ and the latter by a basis of size $\mathsf{DPSP}_{k,\ell}(g)$. Also, for any two sets A and B, $\dim(A+B) \leq \dim(A) + \dim(B)$. Hence, Equation 2.14 reduces to:

$$\begin{aligned}\mathsf{DPSP}_{k,\ell}(f+g) &\leq \dim(\pi(\mathbf{x}^{=\ell}.\partial^{=k}(f))) + \dim(\pi(\mathbf{x}^{=\ell}.\partial^{=k}(g))),\\ &= \mathsf{DPSP}_{k,\ell}(f) + \mathsf{DPSP}_{k,\ell}(g).\end{aligned}$$

□

Chapter 3

Analysis of the $\Sigma\Pi\Sigma\Pi$ Circuit Model

3.1 Upper Bound Statement and Proof Outline

As mentioned in Lemma 1.1, we will assume that the parameters s_1 and s_2 are upper bounded by $\frac{Nd}{\log^5 N}$, and the degree parameter $d = \frac{\sqrt{N}}{\log^4 N}$ in the rest of this chapter. For the sake of simplicity, we shall ignore all ceil and floor notations hereafter.

Let C be a depth four circuit and let $\mathsf{DPSP}_{k,\ell}(C)$ denote the *Dimension of Projected Shifted Partial Derivatives* measure of the polynomial computed by C. We prove the following upper bound on the measure $\mathsf{DPSP}_{k,\ell}(C)$.

Lemma 3.1 *Let C be any depth four circuit computing an N-variate polynomial of degree d, with top fan-in $s \leq N^2$. Let $\tau = 20 \log N$ and $t = \frac{d}{\log^3 N}$ be specific parameters. Then, for all k, ℓ satisfying $\ell < \frac{N}{2} - 2kt$, a random restriction R satisfies the following with high probability:*

$$\mathsf{DPSP}_{k,\ell}(\sigma_R(C)) \leq s \cdot \binom{1 + \frac{\tau.Nd}{\log^5 N.t}}{k} \cdot \binom{N}{\ell + 2kt}.$$

The rest of this chapter is dedicated to proving the above lemma. The main steps involved in the proof are as follows:

- First, we decompose the original circuit C, represented as $C = \sum_i \prod_j P_{ij}$, into two separate depth four circuits, say C_1 and C_2. This is done by splitting every polynomial P_{ij} into two groups of monomials based on some degree criteria (discussed in detail later), and combining all the 'high' degree monomials into the new circuit C_2 by exhaustive multiplication (elaborated later). The circuit C_1 then contains only the remaining 'low' degree monomials.

- Then on the decomposed circuit, we apply the random restriction, derivative, shift and projection operations in that order to evaluate $\mathsf{DPSP}_{k,\ell}(C)$. The degree criteria for the decomposition is chosen in such a manner that in the process of evaluating $\mathsf{DPSP}_{k,\ell}(C)$, the projection and derivative operations together effectively eliminate C_2 from the calculations and allows us to focus on just the restricted degree depth four circuit C_1.

- The circuit C_1 is constructed such that the degree of all the polynomials computed at level 2 sum gates is bounded by a fixed parameter τ (which will be equal to $\Theta(\log N)$). We start by combining these polynomials together into factors of degree at least t (which will be equal to $\frac{d}{\log^3 N}$), which simplifies the estimation of an upper bound on $\mathsf{DPSP}_{k,l}(C_1)$, and hence on $\mathsf{DPSP}_{k,l}(C)$.

The choice of all the parameters used, k, ℓ, t, τ has been reasoned in the final calculations leading to the desired $\tilde{\Omega}(N^{1.5})$ lower bound.

3.2 Using Random Restriction and Projection

Decomposing the circuit. Recall from Equation (1.1), the representation of the circuit C as a sum of products of sparse polynomials.

$$C = T_1 + \ldots + T_s \ ,$$
$$T_i = \prod_{j=1}^{s_1} P_{ij} \quad \forall i \in [s],$$

where Support(P_{ij}) for all i, j is bounded by s_2. We split the polynomials P_{ij}'s into two parts, the first one (say P'_{ij}) containing all those monomials where each variable has individual degree at most two, and the second (say P''_{ij}) comprising of the remaining monomials (i.e at least one of the variables in these monomials has degree at least 3).

Random restriction. Let σ_R be the substitution map that every variable independently survives with probability $p = N^{-\beta} = 1/2$, i.e. for $\beta = 1/\log N$. Here, $R \subseteq [N]$ refers to the set of variables set to zero. Then under the random restriction σ_R, Equation (1.1) reduces to

$$\sigma_R(C(\mathbf{x})) = \sum_{i=1}^{s} \prod_{j=1}^{s_1} \sigma_R(P'_{ij}(\mathbf{x})) + \sigma_R(P''_{ij}(\mathbf{x})). \tag{3.1}$$

After the decomposition, we multiply out all the factors in a term, and split the resulting summands into two groups, one group contributing to the depth four circuit C_1 with all low-

individual-degree terms, and the other group adding into another circuit C_2. More precisely, the term $T_i = \Pi_{j=1}^{s_1} P_{ij}$ is replaced by 2^{s_1} terms, only one of which (the $\prod_{j=1}^{s_1} P'_{ij}$ term) belongs to C_1, and the rest belong to C_2 as follows:

$$\begin{aligned} T_i &= (P'_{i1} + P''_{i1})(P'_{i2} + P''_{i2}) \dots (P'_{is_1} + P''_{is_1}) \\ &= \prod_{j=1}^{s_1} P'_{ij} + \text{ (other terms having at least one } P''_{ij} \text{ factor)} \end{aligned}$$

Similarly manipulating every term T_i of C for $i \in [s]$, from Equation (3.1) we get

$$\sigma_R(C(\mathbf{x})) = \sigma_R(C_1(\mathbf{x}) + \sigma_R(C_2(\mathbf{x})), \tag{3.2}$$

where as explained before, $C_1(\mathbf{x}) = \sum_{i=1}^{s} \prod_{j=1}^{s_1} P'_{ij}(\mathbf{x})$.

Observation 3.2 $\mathsf{DPSP}_{k,\ell}(C_2(\mathbf{x})) = 0$.

Proof: When the partial derivative operation $\partial^{=k}$ (by a multilinear monomial of degree k) is applied on the entire circuit C, the monomials in $C_2(\mathbf{x})$ which have at least one variable with degree three or more, do not reduce to multilinear monomials. Hence, they do not survive after the multilinear projection π. More precisely, $\pi(\partial^{=k}(C_2(\mathbf{x}))) = 0$, and therefore $\pi(\mathbf{x}^{=\ell}.\partial^{=k}(C_2(\mathbf{x})) = 0$. □

As a result of Observation 3.2, C_2 does not contribute to the final DPSP calculation, despite having a much larger size than C_1. Hence we observe that proving a lower bound on the circuit size of C_1 implies a lower bound on the size of the circuit C, since the size of C_1 is at most the size of C. Thus, we focus on lower bounding $C_1(\mathbf{x}) = \sum_{i=1}^{s} \prod_{j=1}^{s_1} P'_{ij}(\mathbf{x})$.

Observation 3.3 *If σ_R is a random restriction chosen as above, then with high probability $\sigma_R(P'_{ij})$ consists of monomials of support at most $\tau' = 10 \log N$ for every $i \in [s], j \in [s_1]$.*

Proof: Consider the random restriction σ_R described earlier, where every variable independently survives with probability $p = N^{-\beta}$, and $R \subseteq [N]$ refers to the set of variables that are set to zero. Under σ_R, a monomial having support at least τ' will survive with probability at most $p^{\tau'}$. The size of Support(P_{ij}) for every polynomial P'_{ij} computed at a level 2 sum gate (Figure 2.1) is bounded by $s_2 \leq \frac{Nd}{\log^5 N}$, and there are total $s.s_1 \leq s.\left(\frac{Nd}{\log^5 N}\right)$ such polynomials, so by union bound the *bad* probability that a monomial having support at least τ' survives in

any of the P'_{ij}'s is at most $s.\left(\frac{Nd}{\log^5 N}\right)^2 p^{\tau'}$.

As mentioned earlier in Section 1.4, the hard polynomial f has degree $d = \frac{\sqrt{N}}{\log^4 N}$. Substituting d and $p = N^{-\beta}$ in the expression for *bad* probability, we get $s.\left(\frac{N^3}{\mathsf{poly}\log N}\right).N^{-\beta\tau'}$. Moreover, as we are trying to achieve a $\tilde{\Omega}(N^{1.5})$ lower bound on s, it is safe to assume $s \leq N^2$ and therefore it suffices that the product $\beta.\tau'$ be a large enough constant (> 5) for the *bad* probability to be ultimately negligible. In our proof, we choose the values of the parameters as $\tau' = 10\log N$ and $\beta = \frac{1}{\log N}$ ($p = 1/2$) to enforce the above argument and keep the *bad* probability low. Thus, for the above setting of parameters β and τ'

$$s.\left(\frac{N^3}{\mathsf{poly}\log N}\right).N^{-\beta\tau'} = \frac{1}{N^5\mathsf{poly}\log N}$$

Hence, with probability at least $1-\tilde{O}(N^{-5})$, $R \subseteq [N]$ is such that $\sigma_R(P'_{ij})$ consists of monomials of support at most τ' for all i, j. □

Since the individual degrees of each variable in P'_{ij} is at most 2, from Observation 3.3, the degree of all the monomials in $\sigma_R(P'_{ij})$ is bounded by $2\tau' = 20\log N = \tau$ (say). Hereafter, it suffices to account for these bounded degree P'_{ij}'s while estimating $\mathsf{DPSP}_{k,\ell}(C)$. Therefore, we slightly abuse the notation and refer to C_1 as C and P'_{ij}'s simply as P_{ij}'s, and assume all P_{ij}'s to have degree bounded by $\tau = 20\log N$.

3.3 Estimating $\mathsf{DPSP}(C)$ for the restricted circuit

Applying the sub-additivity of the DPSP measure (Lemma 2.8) on Equation (1.1), we get:

$$\mathsf{DPSP}_{k,\ell}(C) \leq s.\mathsf{DPSP}_{k,\ell}(T) \tag{3.3}$$

where T is the representative term with maximum DPSP value out of $T_1, \ldots, T_s$. So, to upper bound $\mathsf{DPSP}_{k,\ell}(C)$, it is sufficient to focus on $\mathsf{DPSP}_{k,\ell}(T)$, where T is of the form $T = \Pi_{j=1}^{s_1} P_j$. Inspired from the techniques used in [KST16], we group the P_j's into disjoint subsets, such that the degree of the product of P_j's in every set is at least t and at most $2t$ (where $t = \frac{d}{\log^3 N}$, so $t \gg \tau$) , except perhaps one remaining P_j factor that could not be grouped with any other factor. Hence,

$$T = Q_1 \ldots Q_m \tag{3.4}$$

where every Q_i for all $i \in [m]$ is a product of one or more P_j's, and all except one have degree in $[t, 2t]$. As the P_j's (after random restriction) are all restricted to degree at most τ, we have

$$(m-1).t \leq \tau.s_1$$
$$m \leq 1 + \tau.s_1/t.$$

From the definition of the DPSP measure, we know that

$$\mathsf{DPSP}_{k,\ell}(T) = \dim(\pi(\mathbf{x}^{=\ell}.\partial^{=k}(Q_1 \dots Q_m)). \tag{3.5}$$

Let us first analyse $\partial^{=k}(Q_1 \dots Q_m)$. The k-th order derivative of the product $Q_1 \dots Q_m$ is obtained by applying the chain rule of derivatives. It is a sum of many terms, each term having at most k of the Q_i's *touched* (i.e affected) by the derivative ($k \leq m$). For example, consider the case of $k = 1$ i.e derivative with respect to a variable:

$$\frac{\partial(Q_1 \dots Q_m)}{\partial x_i} = \sum_{i \in [m]} \frac{\partial Q_i}{\partial x_i} \cdot \prod_{j \in [m] \setminus \{i\}} Q_j$$

As k increases, the number of summands increase but every summand has at least $m - k$ of the Q_i's unaffected by the derivative. Fixing this unaffected set of Q_i's, for every summand in $\partial^{=k}(Q_1 \dots Q_m)$, we denote the remaining set of Q_i's as *touched* by the derivative, and identified by the set $A \subseteq [m]$ of size $|A| = k$. For any such set A, we define

$$d_A := \sum_{i \in A} \deg(Q_i),$$
$$V_A := \mathrm{span}_{\mathbb{F}}(\mathbf{x}^{\leq(\ell + d_A - k)}.\Pi_{i \notin A}\ Q_i),$$

where $\mathbf{x}^{\leq(\ell+d_A-k)}$ refers to the set of monomials of degree at most $\ell + d_A - k$. Since the total degrees of the factors Q_i are bounded by $2t$, $d_A \leq 2kt$ for all $A \in \binom{[m]}{k}$. We now make the following observation that is critical to upper bound $\mathsf{DPSP}_{k,\ell}(C)$, which is later used to obtain the final upper bound on $\mathsf{DPSP}_{k,\ell}(C)$.

Observation 3.4 $\mathbf{x}^{=\ell}.\partial^{=k}(T) \subseteq \mathrm{span}_{\mathbb{F}}\{V_A :\ A \in \binom{[m]}{k}\}$.

Proof: As discussed earlier, $\partial^{=k}(T)$ can be split into $\binom{m}{k}$ terms based on the subset A *touched* by the derivative. Then, all these polynomials are multiplied by the shift $\mathbf{x}^{=\ell}$ as follows:

$$\partial^{=k}(T) \subseteq \mathrm{span}_{\mathbb{F}}\{(\partial^{=k}(\Pi_{i \in A}\ Q_i))(\Pi_{i \notin A}\ Q_i) :\ A \in \binom{[m]}{k}\},$$

$$\mathbf{x}^{=\ell}.\partial^{=k}(T) \subseteq \text{span}_{\mathbb{F}}\{(\mathbf{x}^{=\ell}.\partial^{=k}(\Pi_{i\in A}\ Q_i))(\Pi_{i\notin A}\ Q_i) :\ A \in \binom{[m]}{k}\}.$$

As the degree of $\partial^{=k}(\Pi_{i\in A}\ Q_i)$ is at most $d_A - k$,

$$\mathbf{x}^{=\ell}.\partial^{=k}(T) \subseteq \text{span}_{\mathbb{F}}\{(\mathbf{x}^{\leq(\ell+d_A-k)}.\Pi_{i\notin A}\ Q_i) :\ A \in \binom{[m]}{k}\}.$$

Thus, from the definition of the set V_A, the proposition is proved. □

Applying the above proposition, along with the sub-additivity of DPSP (Lemma 2.8) and Observation 2.7, Equation (3.5) can be written as

$$\mathsf{DPSP}_{k,\ell}(T) \leq \sum_{A\in\binom{[m]}{k}} \dim(\pi(V_A)). \tag{3.6}$$

Suppose $\dim(\pi(V_A)) \leq u$, where $u \in \mathbb{N}$. Since $m \leq 1 + \tau.s_1/t$, Equation 3.3 reduces to the following:

$$\mathsf{DPSP}_{k,\ell}(C) \leq s.\binom{1+\tau.s_1/t}{k}.u \tag{3.7}$$

Observation 3.5 $\dim(\pi(V_A)) \leq \binom{N}{\ell+2kt}$ *for every* $A \in \binom{[m]}{k}$.

Proof: Consider any representative V_A for some $A \in \binom{[m]}{k}$. The generators of V_A would consist of polynomials of the form $g(\mathbf{x}) = m(\mathbf{x}).(\Pi_{i\notin A}Q_i)$, where $m(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ is a monomial of degree at most $\ell + d_A - k$, and without loss of generality we can assume that it is a multilinear monomial (otherwise it would be eliminated by π). So, let us denote $m(\mathbf{x})$ by $\mathbf{x}_S$, where S is the set of variables ($S \subseteq [N]$), such that $S = \text{Supp}(m(\mathbf{x}))$.

The number of ways of choosing S such that $\mathbf{x}_S$ is a monomial of degree $\ell + d_A - k$ is given by $\binom{N}{\ell+d_A-k}$. Also, notice that once the set S is fixed, it would also fix the variables that cannot occur in the $(\Pi_{i\notin A}Q_i)$ part of the polynomial $g(\mathbf{x})$, in order to maintain the multilinear property of the projected polynomial $\pi(g(\mathbf{x}))$. Hence, to sum up:

$$\pi(V_A) \subseteq \text{span}_{\mathbb{F}}\left(\{\mathbf{x}_S.\pi(\sigma_S(\Pi_{i\notin A}Q_i)) : S \subseteq [N], |S| = \ell + d_A - k\}\right)$$

$$\Rightarrow \dim(\pi(V_A)) \leq \binom{N}{\ell + d_A - k} \tag{3.8}$$

using $d_A \leq 2kt$ and $\ell + 2kt < N/2$ (proved below),

$$\dim(\pi(V_A)) \leq \binom{N}{\ell + 2kt}. \tag{3.9}$$

□

Plugging this upper bound $u = \binom{N}{\ell+2kt}$ from Observation 3.5 in Equation (3.7), we get

$$\mathsf{DPSP}_{k,\ell}(C) \leq s \cdot \binom{1 + \frac{\tau.s_1}{t}}{k} \cdot \binom{N}{\ell + 2kt},$$

where C is restricted by the random restriction σ_R and $\tau = 20 \log N$ and $s_1 \leq \frac{Nd}{\log^5 N}$. This completes the proof of Lemma 3.1.

Choice of ℓ. In Lemma 3.1, we state the result under the constraint that $\ell + 2kt < N/2$, which was essential to the upper bound in Equation 3.9. We show that this is indeed true for our choice of parameters which is as follows:

$$\begin{aligned}
d &= \frac{\sqrt{N}}{\log^4 N} \\
t &= \frac{d}{\log^3 N} \\
k &= \frac{\delta.d}{t} \\
\delta &= \frac{1}{\log N} \\
\ell &= \frac{N}{2}.\left(1 - \frac{N^{\delta/t} - 1}{N^{\delta/t} + 1}\right) = \frac{N}{N^{\delta/t} + 1}
\end{aligned}$$

Claim 3.6 *for the above choice of parameters, $\ell < N/2 - 2kt$.*

Proof: We prove that the following ratio is greater than 1.

$$\begin{aligned}
\frac{\frac{N}{2} - 2kt}{\ell} &= \frac{(\frac{N}{2} - 2\delta d).(N^{\delta/t} + 1)}{N} \\
&= \left(\frac{1}{2} - \frac{2\delta d}{N}\right).(N^{\delta/t} + 1)
\end{aligned}$$

Thus, we have to prove the following:

$$\begin{aligned}
&\frac{1}{\frac{1}{2}-\frac{2\delta d}{N}} < N^{\delta/t}+1 \\
\Rightarrow &\frac{1}{\frac{1}{2}-\frac{2\delta d}{N}} - 1 < N^{\delta/t} \\
\Rightarrow &\frac{1+\frac{4\delta d}{N}}{1-\frac{4\delta d}{N}} < 2^{1/t}
\end{aligned}$$

As $\frac{4\delta d}{N} \ll 1$ and $e^x \approx 1+x$ for $x \ll 1$,

$$\begin{aligned}
e^{\frac{8\delta d}{N}} &< 2^{1/t} \\
e^{\frac{8\delta dt}{N}} &< 2
\end{aligned}$$

The above is true if $\frac{c_0\delta dt}{N} < 1$ for some constant c_0.

$$\frac{c_0\delta dt}{N} = \frac{c_0.d^2}{N\log^4 N} = \frac{c_0}{\mathsf{poly}\log N}$$

□

Chapter 4

Explicit polynomial of high measure: Proof of Theorem 1

The analysis in the chapter closely follows the analysis in [KLSS14] and [KS15], but we reproduce some of the arguments and calculations here as our choices of parameters (in particular, ℓ, k, t, d, β) are slightly different and suited to our purpose.

4.1 The Hard polynomial family f

As mentioned earlier in Chapter 1, the hard polynomial family is a variant of the Nisan-Wigderson polynomial family (Equation 2.5). The N-th member of this family has degree d and $N = dq$ variables, is parametrized by a number r (fixed later in the analysis), and is defined as follows:

$$\mathrm{NW}_r(x_{1,1}, x_{1,2}, \ldots, x_{d,q}) := \sum_{\substack{h(z)\in\mathbb{F}_q[z] \\ deg(h)\leq r}} \prod_{i\in[d]} x_{i,h(i)} \tag{4.1}$$

where q is chosen as the smallest prime number greater than $d^{1+\alpha}$, where $\alpha < 1$ is a parameter fixed later to achieve the required result. Hence the number of input variables N is such that $d^{2+\alpha} \leq N \leq 2d^{2+\alpha}$.

As part of our proof, we focus on the polynomial f obtained by applying the random restriction described in Chapter 3. Hence, $f = \sigma_R(\mathrm{NW}_r)$ is the hard polynomial for which we prove the lower bound as in Lemma 4.1, borrowing the main ideas from [KS15], but with a slightly different polynomial family $\{f_N\}$, and an updated setting of the basic parameters. In the following lemma, $\beta = \frac{1}{\log N}$, α is as described above, and $t = \frac{d}{\log^3 N}, p = N^{-\beta}$ are as chosen in the circuit upper bound in Chapter 3.

Lemma 4.1 *Let $k = \delta.\frac{d}{t}$ where $0 < \delta < 1$ is a small constant, r be some fixed function dependent on α, β, d, and $\ell = \frac{N}{N^{\delta/t}+1}$. Then with probability at least $1 - \frac{1}{d^{O(1)}}$,*

$$\mathsf{DPSP}_{k,\ell}(f) \geq \frac{1}{d^3}.\min\left(\frac{p^k}{4^k}.\binom{N}{k}.\binom{N}{\ell}, \binom{N}{\ell+d-k}\right).$$

The parameters: The DPSP measure was first used in [KLSS14], where the result was obtained by setting the parameters as $\ell = \frac{N}{2}.(1 - \frac{k \log d}{d})$ and $k = \delta.\frac{d}{t}$, where t was set to be $O(\sqrt{d})$ and $d = N^{1/3}$. The parameters in our proof are set as follows:

$$\begin{aligned}
\ell &= \frac{N}{N^{\delta/t}+1} \\
d &= \frac{\sqrt{N}}{\log^4 N} \\
k &= \delta.\frac{d}{t} \\
t &= d/\log^3 N \\
\delta &= \frac{1}{\log N} \\
r &= \frac{(\alpha+\beta)}{2} \cdot d - 1
\end{aligned}$$

Also, from the inequality $d^{2+\alpha} \leq N \leq 2d^{2+\alpha}$ and $d = \frac{\sqrt{N}}{\log^4 N}$, we find $\alpha = \Theta(\frac{\log\log N}{\log N})$. It must be noted that we make use of few results from [KLSS14] and [KS15], where the choice of parameters were different, but we have verified that those results hold under our choice of parameters.

We approach the proof of Lemma 4.1 in three steps. First, we equate the measure $\mathsf{DPSP}_{k,\ell}(f)$ to the rank of a coefficient matrix M, that exactly captures the effect of *shift* and *derivative* operations on the NW_r polynomial. Second, we construct a square matrix $B = M^T \cdot M$ and recall the notion of *Surrogate Rank* (denoted SurRank) from [KS15] and hence, further reduce our lower bound candidate from $\mathrm{rank}(M)$ to $\mathrm{SurRank}(B)$. Thus, as a consequence of the first two steps:

$$\mathsf{DPSP}_{k,\ell}(f) = \mathrm{rank}(M) \geq \mathrm{SurRank}(B).$$

In the final step, we prove a lower bound on $\mathrm{SurRank}(B)$.

An important aspect of our proof, is the choice of parameters k, ℓ, α, β that constitute the primary difference in our work from [KS15]. Both [KLSS14] and [KS15] use members of the

Nisan-Wigderson family as their choice of hard polynomials, but operate on different underlying arithmetic circuit models. [KLSS14] proves lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi$ circuits, where as [KS15] operates on $\Sigma\Pi\Sigma$ circuits with bottom fan-in restricted by τ. As described in Chapter 3, we too have applied the appropriate random restriction, such that it suffices to assume that all polynomials computed at level 2 gates in our $\Sigma\Pi\Sigma\Pi$ circuit are degree-bounded by $\tau = \Theta(\log N)$. Therefore, the parameter τ used to bound the degree (or support) of intermediate polynomials, forms a common node of importance in [KLSS14],[KS15], and in our proof.

4.2 Proof of Lemma 4.1

The set of multilinear N-variate d-degree monomials is in 1-1 correspondence with $\binom{[N]}{d} = \binom{[d]\times[q]}{d}$. Hence in the arguments below, we naturally associate elements of $\binom{[N]}{d}$ with N-variate d-degree monomials. A monomial in Support(f) corresponds to a univariate polynomial $h \in \mathbb{F}_q[z]$ of degree at most r. As any two different univariate r-degree polynomials over $\mathbb{F}_q$ $(r < q)$ can have at most r roots over any field, we make the following observation.

Observation 4.2 *For two distinct sets $D_1, D_2 \in \binom{[d]\times[q]}{d}$, corresponding to monomials in* Support(f)*, the following holds:*

$$|D_1 \cap D_2| \leq r.$$

Applying the DPSP measure to the polynomial in Equation (4.1) gives us

$$\mathsf{DPSP}_{k,\ell}(f) = \dim\left(\left\{\mathbf{x}_A.\sigma_A(\partial^C f) : A \in \binom{[N]}{\ell}, C \in \binom{[N]}{k}\right\}\right), \tag{4.2}$$

where A and C are the subsets corresponding to the shift and the derivative monomials respectively and $\partial^C(f)$ denotes the partial derivative of f with respect to the monomial $\mathbf{x}_C$. We show that $\mathsf{DPSP}_{k,\ell}(f)$ is equal to the rank of the matrix $M(f)$ as constructed below. In the arguments below, $d' = d - k$.

Construction of the matrix $M(f)$: Define the matrix $M(f)$ as follows:

- The rows of $M(f)$ are labelled by pairs of subsets $(A, C) \in \binom{[N]}{\ell} \times \binom{[N]}{k}$, and
- The columns of $M(f)$ are labelled by subsets $S \in \binom{[N]}{\ell+d'}$.

Each row corresponds to the polynomial $f_{(A,C)} := \mathbf{x}_A.\sigma_A(\partial^C f)$, and the S-th entry of that row is the coefficient of the monomial $\mathbf{x}_S$ in the polynomial $f_{(A,C)}$. Then it is easy to observe the following.

Observation 4.3 $\mathsf{DPSP}_{k,\ell}(f) = \text{rank}(M(f))$.

Hereafter in the arguments, we will refer to $M(f)$ simply as the matrix M. We wish to lower bound $\text{rank}(M)$. First we make the following observation.

Observation 4.4 *M is a 0-1 valued matrix.*

Let A, C, S be subsets of $[N]$, such that $A \subseteq S$ and $(S \setminus A) \cap C = \phi$. Taking cue from the above discussion, we label the $((A, C), S)$-th entry with the set $D \in \binom{[N]}{d}$ computed as $D = (S \setminus A) \uplus C$. We call D a *valid set* if the monomial $\mathbf{x}_D \in \text{Support}(f)$, else we call it an *invalid* set. Instead, if $A \nsubseteq S$ or $(S \setminus A) \cap C \neq \phi$, we simply call the label *not defined*. Observe that an entry in M is non-zero if and only if it is labelled by a valid set. We lower bound the rank of M by calculating a lower bound on the *surrogate rank* of M defined as follows.

Definition 4.5 (Surrogate Rank) *For any matrix $M \in \cup_{m,n\in\mathbb{N}}\mathcal{M}_{m,n}(\mathbb{R})$ where $\mathcal{M}_{m,n}(\mathbb{R})$ denotes the set of real valued $m \times n$ matrices, we define the real symmetric matrix $B := M^T \cdot M$. From the definition of B, it is easy to show that*

$$\text{rank}(M) \geq \text{rank}(B)$$

where the equality holds over the field of real numbers $\mathbb{R}$. Further, by an application of Cauchy-Schwartz on the non-zero eigenvalues of B, [Alo09] obtained the following bound over $\mathbb{R}$:

$$\text{rank}(B) \geq \frac{\text{Tr}(B)^2}{\text{Tr}(B^2)}. \tag{4.3}$$

The above ratio is called the surrogate rank of B, denoted $\text{SurRank}(B)$.

The notion of SurRank has been previously used in [KLSS14] and [KS15] to prove lower bounds. The idea is to exploit the structure of the matrix M, to compute a lower bound on the surrogate rank of B where $B = M^T \cdot M$. Observe that M is a relatively sparse 0-1 matrix. Hence, it becomes simpler to estimate $\text{Tr}(B)$ and $\text{Tr}(B^2)$. The rest of the proof shall be devoted to finding a tight lower bound on $\text{SurRank}(B)$ which would (from Observation 4.3 and Equation 4.3) imply the $\mathsf{DPSP}_{k,\ell}(f)$ lower bound as claimed in Lemma 4.1.

4.3 Lower bounding SurRank(B)

To compute a lower bound on SurRank(B), we lower bound $(\mathrm{Tr}(B))^2$, and upper bound $\mathrm{Tr}(B^2)$. Equation 4.3 thus enables us to lower bound SurRank(B). In the following proof calculations, we use an upper bound on the quantity $H_r(d, w)$ that denotes the number of univariate polynomials in $\mathbb{F}_q[z]$ of degree at most r, having exactly w distinct roots , where $w \in [d]$.

An upper bound on $H_r(d, w)$: A polynomial $h(z) \in \mathbb{F}_q[z]$ of degree at most r, with exactly w distinct roots in $[d]$ must be of the form:

$$h(z) = (z - \alpha_1).(z - \alpha_2) \dots (z - \alpha_w).\hat{h}(z)$$

where $\alpha_i \in [d]$ for $i \in [w]$ and $\hat{h}(z) \in \mathbb{F}_q[z]$ is of degree at most $(r - w)$. Thus, we have

$$H_r(d, w) \leq q^{r-w+1} \cdot \binom{d}{w} \leq q^{r+1} \cdot \left(\frac{d}{q}\right)^w \cdot \frac{1}{w!}. \tag{4.4}$$

4.3.1 Estimating a lower bound on Tr(B)

Since M is a 0-1 valued matrix (Observation 4.4), $\mathrm{Tr}(B) = \mathrm{Tr}(M^T \cdot M)$ is equal to the number of non-zero entries in the matrix M. Hence, $\mathrm{Tr}(B)$ is equal to the number of cells labelled by a valid set. Recall that a set $D \in \binom{[N]}{d}$ labelling a cell in M, is a valid set if $\mathbf{x}_D \in \mathrm{Support}(f)$.

To estimate the number of cells in M labelled by a valid set, we first count all possible valid sets i.e the size of the Support(f), and then multiply this to the number of possible entries in M that can be labelled by a particular fixed valid set. Firstly, it is easy to observe the following:

Observation 4.6 *A set $D \in \binom{[N]}{d}$ labels at least $\binom{d}{k}.\binom{N-d}{\ell}$ entries of M.*

Proof: A set $D \in \binom{[N]}{d}$ labels the $((A, C), S)$-th entry of M if and only if the monomial $\mathbf{x}_S$ is obtained by removing the variables of $\mathbf{x}_C$ and adding the variables of $\mathbf{x}_A$ to the variables in $\mathbf{x}_D$. Hence the number of entries in M labelled by D equals the number of ways we can choose C and A. We can choose the set C in exactly $\binom{d}{k}$ ways, and choose the set A in at least $\binom{N-d}{\ell}$ ways. Thus, a set $D \in \binom{[N]}{d}$ labels at least $\binom{d}{k}.\binom{N-d}{\ell}$ entries of M. □

The size of Support(f) is equal to the number of monomials in NW_r that survive after the

random restriction R. More precisely,

$$f = \sigma_R(\mathrm{NW}_r) = \sum_{D \in \mathrm{Supp}(\mathrm{NW}_r)} e_D.\mathbf{x}_D,$$

where e_D is an indicator random variable which equals 1 if and only if $\sigma_R(\mathbf{x}_D) \neq 0$, i.e the monomial $\mathbf{x}_D$ has no variable in common with the random restriction set R. Let $\mu(f)$ be a random variable that denotes the number of monomials in f, equal to the size of Support(f). We make the following observation:

Observation 4.7 $\mathcal{E}[\mu(f)] = p^d.q^{r+1}$.

Proof: We know that the number of monomials that survive after random restriction R, is equal to

$$\mu(f) = \sum_{D \in \mathrm{Supp}(\mathrm{NW}_r)} e_D$$
$$\Rightarrow \mathcal{E}[\mu(f)] = \sum_{D \in \mathrm{Supp}(\mathrm{NW}_r)} \mathcal{E}[e_D].$$

Since there are q^{r+1} monomials in NW_r all of degree d,

$$\mathcal{E}[\mu(f)] = p^d.q^{r+1}.$$

□

Let $\gamma := p^d.q^{r+1}$. Hence, from Observations 4.6 and 4.7,

$$\mathcal{E}[\mathrm{Tr}(B)] \geq \gamma.\binom{d}{k}.\binom{N-d}{\ell}.$$

The following result has been proved in [KS15] using variance calculation and Chebyshev's inequalities.

Proposition 4.8 ([KS15]) $\Pr[\mathrm{Tr}(B)] \leq \frac{1}{2}.\gamma.\binom{d}{k}.\binom{N-d}{\ell}] \leq \frac{10}{pd^{\alpha}}$.

4.3.2 Estimating an upper bound on $\mathrm{Tr}(B^2)$

From the definition of B, we have $B^2 = (M^T \cdot M)(M^T \cdot M)$. Hence,

$$\mathrm{Tr}(B^2) = \sum_{(A_1,C_1),(A_2,C_2)\in\left(\binom{N}{\ell}\times\binom{N}{k}\right)^2} \; \sum_{S_1,S_2\in\left(\binom{N}{\ell+d'}\right)^2} M_{(A_1,C_1),S_1} \cdot M_{(A_1,C_1),S_2} \cdot M_{(A_2,C_2),S_1} \cdot M_{(A_2,C_2),S_2} \tag{4.5}$$

Since M is a 0-1 matrix (Observation 4.4), $\mathrm{Tr}(B^2)$ is exactly equal to the number of 4-tuples $\{(A_1, C_1), (A_2, C_2), S_1, S_2\}$ such that the four corresponding matrix entries (from the above equation (4.5)) are non-zero. More formally, for any pair of row indices $(A_1, C_1), (A_2, C_2)$ and any pair of column indices S_1, S_2, we define the box $\mathbf{b}$ consisting of the corresponding entries from the matrix M.

$$\begin{aligned}\mathbf{b} :&= \mathrm{box}((A_1, C_1), (A_2, C_2), S_1, S_2)\\ &= (M_{(A_1,C_1),S_1}, M_{(A_1,C_1),S_2}, M_{(A_2,C_2),S_1}, M_{(A_2,C_2),S_2})\end{aligned}$$

Therefore, from equation (4.5), $\mathrm{Tr}(B^2)$ is equal to the number of boxes $\mathbf{b}$ with all four cell-entries non-zero. In other words, it is equal to the number of boxes $\mathbf{b}$ with all four cell entries labelled by valid sets. We formally make a note of this in the following observation.

Observation 4.9 *$\mathrm{Tr}(B^2)$ is equal to the number of boxes $\mathbf{b} = box((A_1, C_1), (A_2, C_2), S_1, S_2)$, such that all four entries in $\mathbf{b}$ are labelled by valid sets. We call such boxes as valid boxes.*

Let $\mathbf{b}$ be a valid box, and D_1, D_2, D_3, D_4 be the labels of the corresponding entries of M, $M_{(A_1,C_1),S_1}, M_{(A_1,C_1),S_2}, M_{(A_2,C_2),S_1}, M_{(A_2,C_2),S_2}$ respectively. To compute an upper bound on the number of such valid boxes, we analyse the structure of the sets D_1, D_2, D_3, D_4 corresponding to the valid boxes. We introduce the following operations on sets, which would bring clarity to the subsequent calculations. For any 2 sets A and B,

- *Subset difference*: $A \setminus\!\setminus B$ equals $A \setminus B$ if $B \subseteq A$, else the operation outputs *not defined*.
- *Disjoint set union*: $A \uplus B$ equals $A \cup B$ if $B \cap A = \phi$ else the operation outputs *not defined*.

Thus the label of the $((A, C), S)$-th cell in M is $(S \setminus\!\setminus A) \uplus C$. Here, $(S \setminus\!\setminus A) \uplus C$ is *not defined*, if either $S \setminus\!\setminus A$ is *not defined* or $(S \setminus\!\setminus A) \cap C \neq \phi$. Similarly, if any entry in the (A, C)-th row is labelled by a set $D \in \binom{[N]}{d}$, then it corresponds to the column identified by $S = (D \setminus\!\setminus C) \uplus A$.

For any box $\mathbf{b} = \text{box}((A_1, C_1), (A_2, C_2), S_1, S_2)$, we define the following sets, as subsets of the sets D_1, D_2, D_3, D_4 labelling the corresponding entries of $\mathbf{b}$:

$E_1 := A_1 \setminus (A_1 \cap A_2)$ $\quad$ $E_2 := A_2 \setminus (A_1 \cap A_2)$

$E_3 := C_1$ $\quad$ $E_4 := C_2$

$E_5 := D_1 \setminus (E_2 \uplus E_3) = D_3 \setminus (E_1 \uplus E_4)$ $\quad$ $E_6 := D_2 \setminus (E_2 \uplus E_3) = D_4 \setminus (E_1 \uplus E_4)$

Observation 4.10 *Based on the sets defined above we make the following observations:*

1. *The set $E_2 \uplus E_3$ is a subset of both D_1 and D_2.*
2. *The set $E_1 \uplus E_4$ is a subset of both D_3 and D_4.*
3. *$D_1 \setminus (E_2 \uplus E_3) = D_3 \setminus (E_1 \uplus E_4)$.*
4. *$D_2 \setminus (E_2 \uplus E_3) = D_4 \setminus (E_1 \uplus E_4)$.*

Proof: Since D_1, D_2, D_3, D_4 are all valid sets, $A_1, A_2 \subseteq S_1$ and $A_1, A_2 \subseteq S_2$. Further, $C_1 \cap (S_1 \setminus A_1) = C_1 \cap (S_2 \setminus A_1) = \phi$. Hence $E_2 \cap E_3 = \phi$, and $E_2 \uplus E_3 \subseteq D_1$. Similarly, $E_2 \subseteq (S_2 \setminus\!\setminus A_1)$ implies $E_2 \uplus E_3 \subseteq D_2$. This proves the first observation, and the second can be proved similarly. To prove the third observation, we show that $D_1 \setminus (E_2 \uplus E_3) = S_1 \setminus (A_1 \cup A_2) = D_3 \setminus (E_1 \uplus E_4)$. Since $D_1 = (S_1 \setminus\!\setminus A_1) \uplus C_1$ and $E_3 = C_2$, the set $D_1 \setminus (E_2 \uplus E_3) = (S_1 \setminus A_1) \setminus E_2$. Thus it is easy to see that $D_1 \setminus (E_2 \uplus E_3) \subseteq S_1 \setminus (A_1 \cup A_2)$. Moreover, as $C_1 \cap (S_1 \setminus A_1) = \phi$, it implies that $S_1 \setminus (A_1 \cup A_2) \subseteq D_1 \setminus (E_2 \uplus E_3)$. Similarly, we can prove that $D_3 \setminus (E_1 \uplus E_4) = S_1 \setminus (A_1 \cup A_2)$. The fourth observation can be proved similarly. □

Finally, it can be verified that the sets D_1, D_2, D_3, D_4 can be expressed as decompositions of the sets $E_1, E_2, E_3, E_4, E_5, E_6$ as stated in the following observation.

Observation 4.11 *The decomposition expressions are as follows:*

1. $D_1 = E_2 \uplus E_3 \uplus E_5$,
2. $D_2 = E_2 \uplus E_3 \uplus E_6$,
3. $D_3 = E_1 \uplus E_4 \uplus E_5$,
4. $D_4 = E_1 \uplus E_4 \uplus E_6$.

Let $|A_1 \cap A_2| = v$, then

$$|E_1| = |E_2| = \ell - v \tag{4.6}$$
$$|E_3| = |E_4| = k \tag{4.7}$$
$$|E_5| = |E_6| = d - (\ell - v + k) \tag{4.8}$$

Proposition 4.12 *Among the sets D_1, D_2, D_3, D_4, only the following scenarios are possible:*

- $D_1 = D_2 = D_3 = D_4$,
- $D_i \setminus D_j \neq \phi$ *for all* $i \neq j$ *(i.e all four sets D_1, D_2, D_3, D_4 are distinct),*
- $D_1 = D_2$ *and* $D_3 = D_4$,
- $D_1 = D_3$ *and* $D_2 = D_4$.

Further, if D_1, D_2, D_3 are all distinct then $\ell - v + k \leq r$ and $d - (\ell - v + k) \leq r$.

Proof: We prove this by considering different possible cases:

- **Case 1**: If $D_1 = D_2$, from Observation 4.11, $E_5 = E_6$ and thus, $D_3 = D_4$.
- **Case 2**: If $D_1 = D_3$, then $E_2 \uplus E_3 = E_1 \uplus E_4$, which implies that $D_2 = D_4$.
- **Case 3**: If $D_1 = D_4$, then from Observation 4.11 , $E_6 \subseteq D_1$ which implies $D_2 \subseteq D_1$. Since $|D_1| = |D_2| = d$, $D_1 = D_2$. Hence $D_1 = D_2 = D_3 = D_4$.

Thus, if D_1 equals any of D_2, D_3, D_4, the proposition holds true. The arguments for the other cases are similar. Suppose D_1, D_2, D_3 are distinct sets. Then, $|D_1 \cap D_2| \geq |E_2 \uplus E_3| = \ell - v + k$. But, from Observation 4.2, if $D_1 \neq D_2$ then $|D_1 \cap D_2| \leq r$. Hence, $\ell - v + k \leq r$. Similarly, since $|D_1 \cap D_3| \geq |E_5| = d - (\ell - v + k)$, it follows that $d - (\ell - v + k) \leq r$. □

We classify the valid boxes based on the four possible scenarios mentioned in Proposition 4.12, and count each of these cases separately. Let $D_1, D_2, D_3, D_4 \in \binom{[N]}{d}$ be the labels of a valid box $\mathbf{b}$ as stated earlier. Then $\mathbf{b}$ belongs to one of the sets as defined below:

$$\begin{aligned}
B_0(D_1) &:= \{\text{box } \mathbf{b} : \text{all four labels equal to } D_1\}, \\
B_1(D_1, D_2) &:= \{\text{box } \mathbf{b} : D_1 = D_3, D_2 = D_4\}, \\
B_2(D_1, D_3) &:= \{\text{box } \mathbf{b} : D_1 = D_2, D_3 = D_4\}, \\
B_3(D_1, D_2, D_3, D_4) &:= \{\text{box } \mathbf{b} : \text{all labels are distinct}\}.
\end{aligned}$$

We further define random variables T_0, T_1, T_2, T_3 as follows, where as previously, e_D is an indicator variable, and equals 1 if x_D survives after the random restriction else 0.

$$T_0 := \sum_{D_1 \in \text{Support}(\text{NW}_r)} e_{D_1}.|B_0(D_1)|, \tag{4.9}$$

$$T_1 := \sum_{D_1 \neq D_2 \in \text{Support}(\text{NW}_r)} e_{D_1}.e_{D_2}.|B_1(D_1, D_2)|, \tag{4.10}$$

$$T_2 := \sum_{D_1 \neq D_3 \in \text{Support}(\text{NW}_r)} e_{D_1}.e_{D_3}.|B_2(D_1, D_3)|, \tag{4.11}$$

$$T_3 := \sum_{D_1 \neq D_2 \neq D_3 \neq D_4 \in \text{Support}(\text{NW}_r)} e_{D_1}.e_{D_2}.e_{D_3}.e_{D_4}.|B_3(D_1, D_2, D_3, D_4)|. \tag{4.12}$$

Hence, from Observation 4.9 and the above arguments, it is easy to observe the following.

Observation 4.13 $\text{Tr}(B^2) = T_0 + T_1 + T_2 + T_3$.

Thus in the arguments that follow, we find suitable upper bounds on T_0, T_1, T_2, T_3.

4.3.2.1 Upper bound for $\mathcal{E}[T_3]$

T_3 corresponds to the number of valid boxes $\mathbf{b}$ where D_1, D_2, D_3, D_4 are all distinct valid sets. As a result of Proposition 4.12, for a valid box if D_1, D_2, D_3 are distinct, D_4 is also distinct. Hence, we approximate T_3 by first counting the number of possible valid boxes for a particular choice of D_1, D_2, D_3, and then multiplying it to the number of ways of choosing $D_1, D_2, D_3 \in \text{Support}(f)$.

Let D_1, D_2, D_3 be distinct valid sets. The following observation shows that a valid box with first three labels D_1, D_2, D_3 can be uniquely determined by fixing the sets E_2, E_3, E_4 and $A_1 \cap A_2$.

Observation 4.14 *For any fixed distinct valid sets D_1, D_2, D_3, the sets A_1, C_1, A_2, C_2 can be uniquely determined, by fixing the sets E_2, E_3, E_4 and $A_1 \cap A_2$.*

Proof: Suppose we fix the set $A_1 \cap A_2$, fixing $E_2 = A_2 \setminus (A_1 \cap A_2)$ will determine A_2. Further, the sets C_1 and C_2 are directly determined by E_3 and E_4 respectively. From Observation 4.10, E_2, E_3, E_4 determine E_1 and therefore A_1. □

We count the number of unique ways to pick E_2, E_3, E_4 and $A_1 \cap A_2$ for a given D_1, D_2, D_3. The choice of E_3 and E_4 are made in $\binom{d}{k}$ ways each, from (Equation **??**). Further, E_2 can be chosen in at most $\binom{d-k}{\ell-v}$ ways, as $E_2 \uplus E_3 \subseteq D_1$ from Observation 4.10. Finally, the set $A_1 \cap A_2$

is chosen in at most $\binom{N-d+2k}{v}$ ways, as $A_1 \cap A_2$ is disjoint from $(D_1 \cup D_3) \setminus (C_1 \cup C_2)$. Hence, the total number of unique choices are at most

$$\binom{d}{k}.\binom{d}{k}.\binom{d-k}{\ell-v}.\binom{N-d+2k}{v}. \tag{4.13}$$

As $\binom{d-k}{\ell-v} \leq 2^{d-k}$ and $v \leq l < \frac{N-d}{2}$ from proposition 4.12, the number of unique ways to pick E_2, E_3, E_4 and $A_1 \cap A_2$ for a given D_1, D_2, D_3 are at most

$$2^{d-k}.\binom{d}{k}^2.\binom{N-d+2k}{\ell}. \tag{4.14}$$

Let $\rho = 2^{d-k}.\binom{d}{k}^2.\binom{N-d+2k}{\ell}$. Then,

$$T_3 \leq \rho. \sum_{D_1 \neq D_2 \neq D_3 \in \text{Support}(\text{NW}_r)} e_{D_1}.e_{D_2}.e_{D_3}. \tag{4.15}$$

Let $\eta = \sum_{D_1 \neq D_2 \neq D_3 \in \text{Support}(\text{NW}_r)} e_{D_1}.e_{D_2}.e_{D_3}$. We upper bound the expected value of η and therefore T_3 in the following proposition, the proof of which is similar to that in [KS15].

Proposition 4.15 *Let γ be as defined in proposition 4.8. Then*

$$\mathcal{E}[\eta] \leq 4 \cdot \gamma^2 \cdot q^{r+1} \cdot \left(\frac{d}{q}\right)^d.$$

Substituting in Equation 4.15, we get

$$\mathcal{E}[T_3] \leq \frac{4}{2^k} \cdot \left(\frac{2}{d^{(\alpha-\beta)/2}}\right)^d \cdot \gamma^2 \cdot \binom{d}{k}^2 \cdot \binom{N-d+2k}{\ell}.$$

Proof: The upper bound on $\mathcal{E}[\eta]$ has been proved in [KS15], and hence omitted. We now achieve the claimed upper bound on $\mathcal{E}[T_3]$, using the above estimate. From Equation 4.15,

$$\mathcal{E}[T_3] \leq \rho.\mathcal{E}[\eta]$$

From equation 4.14,

$$\mathcal{E}[T_3] \leq 2^{d-k} \cdot \binom{d}{k}^2 \cdot \binom{N-d+2k}{\ell} \cdot 4 \cdot \gamma^2 \cdot q^{r+1} \cdot \left(\frac{d}{q}\right)^d$$

Since $r+1 = \frac{\alpha+\beta}{2(1+\alpha)}.d$ and $q \geq d^{1+\alpha}$,

$$\mathcal{E}[T_3] \leq \frac{4}{2^k} \cdot \left(\frac{2}{d^{(\alpha-\beta)/2}}\right)^d \cdot \gamma^2 \cdot \binom{d}{k}^2 \cdot \binom{N-d+2k}{\ell}.$$

□

Eventually, $\mathcal{E}[T_3]$ is found to be negligible in comparison to $\mathcal{E}[T_0+T_1+T_2]$ and hence does not contribute to the final expected value of $\mathrm{Tr}(B^2)$.

4.3.2.2 Upper bound for $\mathcal{E}[T_0]$

The following observation is used to upper-bound T_0, T_1, T_2 in the subsequent arguments.

Observation 4.16 *For any set $D_1 \in \binom{[N]}{d}$ and any row (A, C) of the matrix M, there can be at most one cell in that row with the label D_1.*

Proof: Suppose there are two cells in the row (A, C) labelled by a set D_1, corresponding to the columns S_1 and S_2 respectively. Then, $D_1 = (S_1 \setminus\setminus A_1) \uplus C_1 = (S_2 \setminus\setminus A_1) \uplus C_1$. This implies $S_1 = S_2$, as $S_1 \setminus A_1 = S_2 \setminus A_1$. □

From Observation 4.16, for all the boxes in $B_0(D_1)$ or $B_2(D_1, D_2)$, the columns S_1 and S_2 must be the same. Keeping the above in mind, we make another important observation about the boxes in $B_0(D_1)$.

Observation 4.17 *For every box $\mathbf{b} \in B_0(D_1)$ defined as $\mathbf{b} = box((A_1, C_1), (A_2, C_2), S_1, S_2)$, $A_1 = A_2$ and $C_1 = C_2$.*

Proof: Given a box $\mathbf{b} \in B_0(D_1)$, we know that all four cells of $\mathbf{b}$ are labelled by a valid set, say D_1. Comparing the entries in the column corresponding to S_1,

$$D_1 = D_3 \tag{4.16}$$

$$\Rightarrow (S_1 \setminus\setminus A_1) \uplus C_1 = (S_1 \setminus\setminus A_2) \uplus C_2. \tag{4.17}$$

From Observation 4.11, $E_1 \subseteq D_3 = D_1$ and $E_1 \subseteq A_1$. But D_1 and A_1 are disjoint sets, hence E_1 must be an empty set. Similarly, $E_2 \subseteq D_1 = D_3$, $E_2 \subseteq A_2$ and $D_3 \cap A_2 = \phi$ together imply that E_2 is an empty set. Substituting E_1, E_2 as empty sets in the expressions for D_1 and D_3 (Observation 4.11), and again using the fact that $D_1 = D_3$, we get $E_3 = E_4$ i.e $C_1 = C_2$.

Plugging in $C_1 = C_2$ in Equation 4.17 also proves that $A_1 = A_2$. □

From Observations 4.16 and 4.17, we prove the following upper bound on $\mathcal{E}[T_0]$:

Proposition 4.18

$$|B_0(D_1)| \leq \binom{N-d+k}{\ell}.\binom{d}{k}, \text{ and hence}$$
$$\mathcal{E}[T_0] \leq \gamma.\binom{N-d+k}{\ell}.\binom{d}{k}.$$

Proof: For any fixed set D_1 of size d, we can pick the set $C_1 \subseteq D_1$ in $\binom{d}{k}$ ways and the set A_1 (disjoint with $D_1 \setminus C_1$) in $\binom{N-d+k}{\ell}$ ways. The expression for $\mathcal{E}[T_0]$ follows straight from definition of T_0, where γ is as defined in Proposition 4.8. □

4.3.2.3 Upper bound for $\mathcal{E}[T_1]$

Let $D_1, D_2 \in \binom{[N]}{d}$ be distinct sets, such that $\mathbf{x}_{D_1}, \mathbf{x}_{D_2} \in \text{Support}(f)$. A valid box $\mathbf{b} = \text{box}((A_1, C_1), (A_2, C_2), S_1, S_2)$ is in $B_1(D_1, D_2)$ if the labels satisfy the following: $D_1 = D_3$ and $D_2 = D_4$. Recall that the proof of Observation 4.17 was also based on the premise that $D_1 = D_3$. Hence from the arguments in the proof of Observation 4.17, $A_1 = A_2 = A$ and $C_1 = C_2 = C$. In Proposition 4.19, $w = |D_1 \cap D_2|$.

Proposition 4.19

$$|B_1(D_1, D_2)| = \binom{N-2d+w+k}{\ell}.\binom{w}{k}, \text{ and hence}$$
$$\mathcal{E}[T_1] \leq d.\frac{\gamma^2}{d^{(\alpha-3\beta)k}.k!}.\binom{N-2d+2k}{\ell}.$$

Proof: We fix two sets $D_1, D_2 \in \binom{[N]}{d}$ and count the number of rows (A, C), such that D_1 and D_2 are the first two labels of the box $\mathbf{b} = \text{box}((A, C), (A, C), S_1, S_2)$. Since $C \subset D_1 \cap D_2$ and $|D_1 \cap D_2| = w$, we can pick C in $\binom{w}{k}$ ways. And for every choice of C, we can pick the set A which is disjoint from $(D_1 \cup D_2) \setminus C$, in $\binom{N-2d+w+k}{\ell}$ ways ($|(D_1 \cup D_2) \setminus C| = 2d - w - k$). Hence,

$$|B_1(D_1, D_2)| = \binom{N-2d+w+k}{\ell}.\binom{w}{k}$$

From Equation 4.10,

$$T_1 = \sum_{D_1 \in \text{Support}(\text{NW}_r)} \sum_{w \geq k} \sum_{\substack{D_2 \in \text{Support}(\text{NW}_r) \\ D_2 \neq D_1, |D_1 \cap D_2| = w}} e_{D_1}.e_{D_2}.|B_1(D_1, D_2)|$$

$$\Rightarrow \mathcal{E}[T_1] = \sum_{D_1 \in \text{Support}(\text{NW}_r)} \sum_{w \geq k} \sum_{\substack{D_2 \in \text{Support}(\text{NW}_r) \\ D_2 \neq D_1, |D_1 \cap D_2| = w}} p^d.p^{d-w}.\binom{N-2d+w+k}{\ell}.\binom{w}{k}$$

$$\leq p^{2d}. \sum_{D_1 \in \text{Support}(\text{NW}_r)} \sum_{w \geq k} H_r(d,w).p^{-w}.\binom{N-2d+w+k}{\ell}.\binom{w}{k}.$$

From equation (4.4):

$$\leq p^{2d}. \sum_{D_1 \in \text{Support}(\text{NW}_r)} \sum_{w \geq k} q^{r+1}.\left(\frac{d}{pq}\right)^w.\frac{1}{w!}.\binom{N-2d+w+k}{\ell}.\binom{w}{k}.$$

Recall that $N = d.q \geq d^{2+\alpha}$ where $\alpha < 1$, and $p = N^{-\beta}$, then

$$\mathcal{E}[T_1] \leq p^{2d}.q^{r+1}. \sum_{D_1 \in \text{Support}(\text{NW}_r)} \sum_{w \geq k} \left(\frac{1}{d^{\alpha-3\beta}}\right)^w.\frac{1}{w!}.\binom{N-2d+w+k}{\ell}.\binom{w}{k}$$

The term $\left(\frac{1}{d^{\alpha-3\beta}}\right)^w.\frac{1}{w!}.\binom{N-2d+w+k}{\ell}.\binom{w}{k}$ attains its maximised value at $w = k$ as $\beta = 1/\log N \ll \alpha = \Theta(\log\log N/\log N)$. Hence,

$$\mathcal{E}[T_1] \leq d.\frac{\gamma^2}{d^{(\alpha-3\beta)k}.k!}.\binom{N-2d+2k}{\ell}.$$

This completes the proof of the proposition. □

4.3.2.4 Upper bound for $\mathcal{E}[T_2]$

From Observation 4.16, any box $\mathbf{b} = \text{box}((A_1, C_1), (A_2, C_2), S_1, S_2)$ in $B_2(D_1, D_3)$ has $S_1 = S_2 = S$. Moreover, $D_1 = D_2 = (S \setminus\setminus A_1) \uplus C_1$ and $D_3 = D_4 = (S \setminus\setminus A_2) \uplus C_2$. Let $u := |C_1 \cap C_2|$ and w as defined in Proposition 4.19, then

Proposition 4.20

$$|B_2(D_1, D_2)| = \sum_{0 \leq u \leq k} \binom{N-2d+w+k}{\ell-d+k+w-u}.\binom{d-w}{k-u}^2.\binom{w}{u}, \text{ and hence}$$

$$\mathcal{E}[T_2] \leq dk.\gamma^2.\binom{N-2d+k}{\ell-d+k}.\binom{d}{k}^2.$$

The proof of this proposition is very similar to the proof of Proposition 4.19, hence omitted. Here, the maxima of the relevant expression is achieved at $w = u = 0$.

4.3.3 Putting it together: Proof of lemma 4.1

From Observation 4.13, we know that $\mathrm{Tr}(B^2) = T_0 + T_1 + T_2 + T_3$. We recall the upper bounds from Propositions 4.18, 4.19, 4.20 and 4.15.

$$\begin{aligned}
\mathcal{E}[T_3] &\leq \frac{4}{2^k} \cdot \left(\frac{2}{d^{(\alpha-\beta)/2}}\right)^d \cdot \gamma^2 \cdot \binom{d}{k}^2 \cdot \binom{N-d+2k}{\ell} \\
\mathcal{E}[T_0] &\leq \gamma.\binom{N-d+k}{\ell}.\binom{d}{k} \\
\mathcal{E}[T_1] &\leq d.\frac{\gamma^2}{d^{(\alpha-3\beta)k}.k!}.\binom{N-2d+2k}{\ell} \\
\mathcal{E}[T_2] &\leq dk.\gamma^2.\binom{N-2d+k}{\ell-d+k}.\binom{d}{k}^2
\end{aligned}$$

Comparing the above equations, it can be observed that the upper bound on $\mathcal{E}[T_2]$ dominates upper bounds on $\mathcal{E}[T_0]$ and $\mathcal{E}[T_3]$. Hence, we assume $T_0, T_3 \leq T_2$ which implies $\mathrm{Tr}(B^2) \leq T_1 + 3T_2$. Thus, we get the following result.

Proposition 4.21 *Using Markov's inequality, with probability at least* $1 - \frac{1}{d}$,

$$\mathrm{Tr}(B^2) \leq d^2 \cdot \frac{\gamma^2}{d^{(\alpha-3\beta)k} \cdot k!} \cdot \binom{N-2d+2k}{\ell} + 3d^2 \cdot k \cdot \gamma^2 \cdot \binom{N-2d+k}{\ell-d+k} \cdot \binom{d}{k}^2.$$

Proof: From Markov's inequality, $\Pr\{T_1 > d.\mathcal{E}[T_1]\} < \frac{1}{d}$ and $\Pr\{T_2 > d.\mathcal{E}[T_2]\} < \frac{1}{d}$. Considering the complimentary event of both, with probability at least $(1 - \frac{1}{d})$, $T_1 \leq d.\mathcal{E}[T_1]$ and $T_2 \leq d.\mathcal{E}[T_2]$. Plugging in the bounds from Propositions 4.19 and 4.20 into the equation $\mathrm{Tr}(B^2) \leq T_1 + 3T_2$, we get the claimed upper bound. □

From Proposition 4.8, $\Pr\{\mathrm{Tr}(B) > \frac{1}{2} \cdot \gamma \cdot \binom{d}{k} \cdot \binom{N-d}{\ell}\}$ is at least $1 - \frac{10}{pd^\alpha}$. Combining this with Proposition 4.21, with probability at least $1 - \frac{1}{d^{O(1)}}$,

$$\begin{aligned}\text{SurRank}(B) &\geq \frac{(\text{Tr}(B))^2}{\text{Tr}(B^2)} \\ &\geq \frac{\frac{1}{4}\cdot\gamma^2\cdot\binom{d}{k}^2\cdot\binom{N-d}{\ell}^2}{d^2\cdot\frac{\gamma^2}{d^{(\alpha-3\beta)k}\cdot k!}\cdot\binom{N-2d+2k}{\ell}+3d^2\cdot k\cdot\gamma^2\cdot\binom{N-2d+k}{\ell-d+k}\cdot\binom{d}{k}^2}.\end{aligned}$$

We split the denominator based on one summand dominating another,

$$\geq \min\left(\frac{\frac{1}{4}\cdot\gamma^2\cdot\binom{d}{k}^2\cdot\binom{N-d}{\ell}^2}{2d^2\cdot\frac{\gamma^2}{d^{(\alpha-3\beta)k}\cdot k!}\cdot\binom{N-2d+2k}{\ell}},\ \frac{\frac{1}{4}\cdot\gamma^2\cdot\binom{d}{k}^2\cdot\binom{N-d}{\ell}^2}{6d^2\cdot k\cdot\gamma^2\cdot\binom{N-2d+k}{\ell-d+k}\cdot\binom{d}{k}^2}\right).$$

The first ratio can be split into two factors and separately lower bounded as follows:

$$d^{\alpha k}\cdot k!\cdot\binom{d}{k}^2\geq\frac{1}{2^k}\cdot\binom{N}{k},\ \text{and}$$
$$\frac{\binom{N-d}{\ell}^2}{\binom{N-2d+2k}{\ell}}\geq\frac{1}{2^k.d^3}\cdot\binom{N}{\ell}.$$

The second ratio is lower bounded as follows:

$$\frac{\binom{N-d}{\ell}^2}{\binom{N-2d+k}{\ell-d+k}}\geq\frac{1}{d^3}\cdot\binom{N}{\ell+d-k}.$$

Hence, the final lower bound expression reduces to

$$\text{SurRank}(B)\geq\frac{1}{d^3}.\min\left(\frac{p^k}{4^k}.\binom{N}{k}.\binom{N}{\ell},\binom{N}{\ell+d-k}\right).$$

where $p=N^{-\beta}$ as chosen earlier. This completes the proof of Lemma 4.1.

4.4 Completing the proof of Theorem 1

Now, we prove the final lower bound result by combining Lemma 4.1 with Lemma 3.1. Thus, $\mathsf{DPSP}_{k,\ell}(\sigma_R(C))=\mathsf{DPSP}_{k,\ell}(f)$ where $f=\sigma_R(\text{NW}_r)$. The random restriction σ_R is guaranteed to exist with high probability from Observation 3.3. Also taking union of the probabilities from Observation 3.3 and Lemma 4.1, the following lower bound is also under the high probability of $1-\frac{1}{N^{O(1)}}$.

$$s \geq \frac{\mathsf{DPSP}_{k,\ell}(NW_r)}{\binom{1+\frac{\tau.s_1}{t}}{k} \cdot \binom{N}{\ell+2kt}} \tag{4.18}$$

$$\geq \frac{1}{d^3} \cdot \min\left(\frac{\left(\frac{p}{4}\right)^k \cdot \binom{N}{k} \cdot \binom{N}{\ell}}{\binom{1+\frac{\tau.s_1}{t}}{k} \cdot \binom{N}{\ell+2kt}}, \frac{\binom{N}{\ell+d-k}}{\binom{1+\frac{\tau.s_1}{t}}{k} \cdot \binom{N}{\ell+2kt}} \right) \tag{4.19}$$

Recall the choice of the parameters (Paragraph 4.1) $t = \frac{d}{\log^3 N}$, $k = \delta\frac{d}{t}$, $\delta = 1/\log N$ and $\ell = \frac{N}{N^{\delta/t}+1}$. Using Lemma 2.10, and substituting $\ell = \frac{N}{N^{\delta/t}+1}$,

$$\begin{aligned}
\frac{\binom{N}{\ell}}{\binom{N}{\ell+2kt}} &= \frac{(\ell+2kt)(\ell+2kt-1)\ldots(\ell+1)}{(N-\ell)(N-\ell-1)\ldots(N-\ell-2kt)} \\
&\approx \left(\frac{N}{\ell}-1\right)^{-2kt} \\
&= (N^{\delta/t})^{-2kt} \\
&= N^{-2\delta.k}.
\end{aligned}$$

Similarly, the second term of the minima is approximated using $\ell \gg (d-k)$ and $k \ll t$,

$$\begin{aligned}
\frac{\binom{N}{\ell+d-k}}{\binom{N}{\ell+2kt}} &\approx \left(\frac{N}{\ell}-1\right)^{d-k-2kt} \\
&= (N^{\delta/t})^{d-k-2kt} \\
&\approx N^{\delta/t.(1-2\delta).d} \\
&= N^{(1-2\delta).k}.
\end{aligned}$$

Substituting in the final lower bound, and approximating $\binom{1+\frac{\tau.s_1}{t}}{k}$ by $\binom{\frac{\tau.s_1}{t}}{k}$:

$$\begin{aligned}
s &\geq \frac{1}{d^3} \cdot \min\left(\left(\frac{p}{4}\right)^k \cdot \frac{\binom{N}{k}}{\binom{\frac{\tau.s_1}{t}}{k}} \cdot N^{-2\delta.k}, \frac{N^{(1-2\delta).k}}{\binom{\frac{\tau.s_1}{t}}{k}} \right) \\
&\geq \min(L_1, L_2),
\end{aligned}$$

where $L_1 = \frac{\left(\frac{p}{4}\right)^k \cdot \binom{N}{k} \cdot N^{-2\delta.k}}{d^3 \cdot \binom{\frac{\tau.s_1}{t}}{k}}$ and $L_2 = \frac{N^{(1-2\delta).k}}{d^3 \cdot \binom{\frac{\tau.s_1}{t}}{k}}$. Observe that $L_1 = \frac{\left(\frac{p}{4}\right)^k \cdot \binom{N}{k}}{N^k} \cdot L_2$, i.e $L_1 \leq L_2$ for all values of N. Therefore, we compute a lower bound on L_1, that will imply the required lower

bound on s.

$$L_1 = \frac{1}{d^3.N^{2\delta.k}} \cdot \left(\frac{p}{4}\right)^k \cdot \frac{\binom{N}{k}}{\binom{\frac{\tau.s_1}{t}}{k}} \tag{4.20}$$

Using $p = N^{-\beta}$, $s_1 \leq \frac{Nd}{\log^5 N}$ and the estimate from Equation 2.9,

$$\geq \frac{1}{d^3} \cdot \left(\frac{\frac{N}{k}}{4.N^{\beta+2\delta} \cdot \frac{e\tau Nd}{kt \log^5 N}}\right)^k \tag{4.21}$$

$$\geq \frac{1}{d^3} \cdot \left(\frac{t.\log^5 N}{4.N^{\beta+2\delta}.e.\tau.d}\right)^k \tag{4.22}$$

Recall that $\beta = \delta = 1/\log N$, $\tau = 20 \log N$. $t = \frac{d}{\log^3 N}$ and $k = \delta.\frac{d}{t}$, for which the above equation equals

$$= \frac{1}{d^3} \cdot \Omega\left(\log N\right)^k \tag{4.23}$$

$$= \omega\left(\frac{Nd}{\log^5 N}\right) \tag{4.24}$$

This completes the proof of Theorem 1.

4.5 Future Work: Possible directions

Following the recent breakthrough result by [KST16], a lot of interest has been generated regarding improved lower bounds for constant depth arithmetic circuits. The proof of an "almost cubic lower bound" for general depth three circuits involves a major step, where 'heavy' product gates are eliminated from the target circuit, i.e gates with total degree higher than a fixed threshold, by considering the circuit modulo a particular set of linear polynomials. This restricts the circuit to an affine subspace, wherein they achieve stronger lower bounds. This technique is inspired from the previous quadratic lower bound result by [SW99].

We too employ a similar technique for depth four circuits, wherein we remove gates with high in-degree and isolate the remaining low-degree gates in the circuit, to achieve stronger lower bound on the total size of $\Sigma\Pi\Sigma\Pi$ circuits. Particularly, in order to achieve a similar situation as in [KST16], we need to consider the field of polynomials taken modulo higher degree (more than 1) polynomials. This might result in a more complex algebraic structure as compared to

affine subspaces in [KST16], and therefore needs to be analysed carefully.

Another direction to build on our result, is to manipulate the choice of ℓ. The current choice is made with the purpose to maximize the value of the minimum of the two numerators in Equation 4.19, $\left(\binom{N}{k} \cdot \binom{N}{\ell}\right)$ and $\binom{N}{\ell+2kt}$, by choosing a ℓ for which both quantities are as close as possible. However, it is conceivable that a different choice of ℓ, that gives an inferior lower bound expression than Equation 4.23, but could result in a stronger depth four lower bound result by allowing a larger choice for d. That would imply a different choice of the hard NW_r polynomial family. Here, we are forced to restrict d to at most $\sqrt{N}$ to enforce the constraint proved in Claim 3.6, that also dictates the choice of parameters k and t. But, it is worth analysing whether this constraint is a hard necessity, and if we can achieve close to a quadratic lower bound (for depth four circuits computing multilinear polynomials) in its absence.

Chapter 5

Previous Lower Bounds for Depth Four Circuits

In this section, we analyse two previous results for general depth four circuits, and compare our techniques with those results in detail. Both the results have related approaches (but different from ours) to arithmetic circuit lower bounds, and both of them deal with general circuits having constant depth. In the rest of this chapter we discuss these results, the main techniques and ideas involved, and their implications to the model of our interest, depth four arithmetic circuits.

5.1 Shoup-Smolensky: Polynomial Evaluation

The first breakthrough result for constant depth arithmetic circuits was achieved by Shoup and Smolensky in 1997. They employed arithmetic circuits to solve the well-known problem of multi-point evaluation of a univariate polynomial. The goal is to evaluate a given univariate n-degree polynomial f, with coefficients in $\mathbb{C}$, on a fixed set of points $p_1, p_2, \ldots, p_n \in \mathbb{C}$. Therefore the arithmetic circuit C designed for the above computation, takes as input the set of coefficients $a_i \in \mathbb{C}$ $(i \in [n])$ of the polynomial $f = a_1 + a_2x + \ldots + a_nx^{n-1}$, and outputs the evaluations $f(p_1), f(p_2), \ldots, f(p_n)$. Hereafter in the analysis, circuits of the above type are referred to as *polynomial evaluation* circuits.

The problem of multi-point polynomial evaluation has been well-studied over the decades for its implications connected to algebraic complexity. In the special case when the points $p_1, \ldots, p_n$ are n-th roots of unity, the problem is known as Discrete Fourier Transform (DFT), which has an arithmetic circuit of size $O(n \log n)$ and depth $O(\log n)$, where n is a power of 2. [SS97]

proved the existence of a set of points $p_1, \ldots, p_n$ such that any arithmetic circuit for polynomial evaluation at these points must have at least superlinear size.

It can be observed that the polynomial evaluation circuit C described above, essentially computes a linear transformation of the coefficients $a_1, \ldots, a_n \in \mathbb{C}$, into the multi-point evaluations $f(p_1), f(p_2), \ldots, f(p_n) \in \mathbb{C}$. The transformation can be represented by the Vandermonde matrix V where the (i,j)-th entry is given by $v_{ij} = p_i^{j-1}$ for all $i, j \in [n]$. Since all the intermediate polynomials computed in C are linear forms (total degree one) in the variables $a_1, a_2, \ldots, a_n$, we focus on a restricted class of arithmetic circuits called as *linear circuits* defined as follows.

Definition 5.1 (Linear Circuits) *A linear circuit over a field $\mathbb{F}$ is an arithmetic circuit, where every gate is either an input gate (leaf node) or an addition gate, i.e there are no multiplication gates. Further, the addition gates have unbounded fan-in and every incoming edge is labelled with a field element. The size and depth of a linear circuit are defined in the same way as for general arithmetic circuits.*

Thus, a linear circuit C with n input nodes and m output nodes, computes a linear transformation $T : \mathbb{F}^n \mapsto \mathbb{F}^m$ defined by the $m \times n$ matrix A_C. The entries of $A_C = \{a_{ij}\}$ are computed as follows: The weight of a path is the product of the weights of all its edges, and a_{ij} is the sum of weights of all paths from the j-th input gate ($j \in [n]$) to the i-th output gate ($i \in [m]$). For infinite fields, [Str73] showed that if a linear transformation T is computed by an arithmetic circuit C', then there is also a *linear circuit* C computing T, with size and depth of C within constant factor of the size and depth of C'. Precisely, if the linear circuit C has size s and depth d, the equivalent general arithmetic circuit has size at most $16s$ and depth at most d. Therefore, the lower bound proved by [SS97] on size of a depth-d linear circuit, implies the same asymptotic lower bound on the size of general depth-d arithmetic circuits.

Let C be a linear circuit. Let the linear transformation computed by C be defined by the square matrix $A \in \mathbb{C}^{n \times n}$, with entries $a_{ij} \in \mathbb{C}$ for all $i, j \in [n]$. Let $L_A(n)$ denote the set of all monomials of degree m in $\{a_{ij}\}$, and let $D_A(n)$ be the dimension of the $\mathbb{Q}$-linear space spanned by the monomials in $L_A(n)$. Then, we have the following result.

Lemma 5.2 ([SS97]) *Let C be a linear circuit of size s and depth d, computing a linear transformation T over $\mathbb{C}^n$. Let $A = A_C$ be the associated matrix. Then for $r = \lceil s/d \rceil$,*

$$D_A(n) \leq \binom{n+r}{n}^d.$$

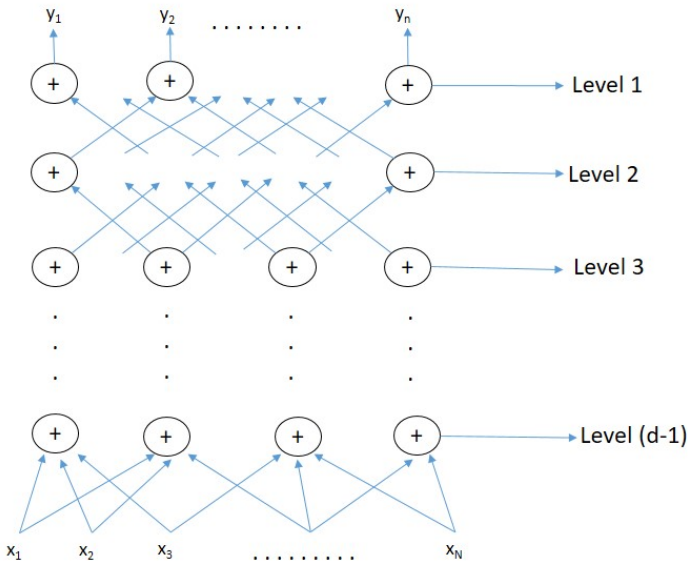

Figure 5.1: Depth-d Linear Circuit

Proof: For this proof, let L and D denote $L_A(n)$ and $D_A(n)$ respectively. Consider the graph of the circuit as depicted in the figure 5.1, with the *level* of each gate as defined for general arithmetic circuits. For $1 \leq i \leq d$, let s_i denote the number of outgoing edges from gates at level i. Hence, the size of the circuit C is equal to $s = \sum_{i \in [d]} s_i$, i.e the total number of edges. Also define W_i as the set of edge weights of the outgoing edges from level i gates. Hence, $|W_i| \leq s_i$ depending on number of distinct edge weights.

Each entry a_{ij} in A can be expressed as a sum of products of the form $(\alpha_1 \dots \alpha_d)$ where $\alpha_i \in W_i \cup \{1\}$ for all $i \in [d]$, as each product corresponds to a path from x_j to y_i. The value 1 for some α_i's takes care of the paths where some levels are skipped. For example if there exists an edge from a level 2 gate to level e ($e > 3$) gate, α_2 is equal to the weight of that edge, and $\alpha_3 = \alpha_4 = \ldots = \alpha_{e-1} = 1$. Further, each element in L is an n-degree monomial in $\{a_{ij}\}$. Suppose $m = a_{i_1 j_1} a_{i_2 j_2} \dots a_{i_n j_n}$ is a monomial in L, where $a_{i_k j_k}$ is a sum of products of the form $(\alpha_1^{(k)} \dots \alpha_d^{(k)})$. Then, m is expressed as a sum of products of the form:

$$(\alpha_1^{(1)} \dots \alpha_1^{(n)}) \dots (\alpha_d^{(1)} \dots \alpha_d^{(n)}) \tag{5.1}$$

where $\alpha_i^{(k)} \in W_i \cup \{1\}$ for all $i \in [d], k \in [n]$. Let Γ be the set of all such products (Equation 5.1), then each element in L belongs to the set $\text{span}_{\mathbb{Z}}(\Gamma)$. Hence D, which is the dimension of the span of L over $\mathbb{Q}$, is at most the cardinality of the set Γ, i.e $D \leq |\Gamma|$.

To calculate the size of the set Γ, we first count the number of products of the form $(\alpha_i^{(1)} \dots \alpha_i^{(n)})$. As $\alpha_i \in W_i$, it can be chosen in at most s_i ways. So, the number of products of the form $(\alpha_i^{(1)} \dots \alpha_i^{(n)})$ are exactly equal to the number of monomials of degree at most n (few of the $\alpha_i^{(k)}$ can be 1) in s_i variables, i.e $\binom{n+s_i}{n}$. Thus, counting for all $i \in [d]$, we get

$$\begin{aligned} D &\leq \prod_{i=1}^{d} \binom{n+s_i}{n} \\ &= \prod_{i=1}^{d} \prod_{j=1}^{n} \frac{n+s_i-j+1}{j} \\ &= \prod_{j=1}^{n} j^{-d} \prod_{i=1}^{d} (n+s_i-j+1). \end{aligned}$$

From arithmetic-geometric mean inequality,

$$\prod_{i=1}^{d} (n+s_i-j+1) \leq (n+\frac{s}{d}+1-j+1)^d.$$

Therefore, substituting $r = \lceil \frac{s}{d} \rceil$,

$$D \leq \prod_{j=1}^{n} \frac{(n+r-j+1)^d}{j^d} = \binom{n+r}{n}^d.$$

□

The above lemma forms the foundation of the main lower bound result from [SS97]. The set of points $p_1, \dots, p_n$ considered by [SS97] are algebraically independent over $\mathbb{Q}$, i.e they do not satisfy any non-trivial polynomial equation over $\mathbb{Q}$. In other words, for every polynomial $f(x_1, \dots, x_n) \in \mathbb{Q}[x_1, \dots, x_n]$ that is not identically zero, $f(p_1, \dots, p_n) \neq 0$.

Theorem 2 ([SS97]) *Let $p_1, p_2, \dots, p_n$ be complex numbers, algebraically independent over $\mathbb{Q}$, with $n > 1$. Any depth d linear circuit for polynomial evaluation at these points, must have size*

$$s > K.dn^{1+1/d}$$

where K is an absolute constant and $d \leq \log n / \log 3$.

Proof: Consider the linear circuit C for polynomial evaluation at the fixed set of points $p_1, p_2, \ldots, p_n \in \mathbb{C}$. Let $o_i = f(p_i)$ be the evaluations of the polynomial $f = a_1 + a_2 x + \ldots + a_n x^{n-1}$ at these points. Thus, the evaluation can be represented as the following transformation by the Vandermonde matrix $V = \{v_{ij}\}_{n \times n}$.

$$\begin{bmatrix} o_1 \\ \vdots \\ o_n \end{bmatrix} = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nn} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

The linear circuit C computes the linear transformation defined by the matrix V. Consider the product of n elements in V, taking one from each row. They are precisely sets of products of the form $(p_1^{e_1} \ldots p_n^{e_n})$ where $e_i \in [0, n-1]$. Hence, there are n^n possible products of the above kind.

Observation 5.3 *All n^n products described above are linearly independent over $\mathbb{Q}$.*

Proof: Note that the p_i's are algebraically independent over $\mathbb{Q}$. This implies that the set $\{p_1^{e_1} \ldots p_n^{e_n} : e_i \in [0, n-1] \text{ for all } i \in [n]\}$ is linearly independent over $\mathbb{Q}$, otherwise the dependence relation would form a non-trivial polynomial equation satisfied $p_1, \ldots, p_n$ which contradicts the algebraic independence assumption. □

Thus, from the above observation and Lemma 5.2,

$$\binom{n+r}{n}^d \geq D_V(n) \geq n^n.$$

Taking logarithm on both sides and applying Stirling's approximation (Equation 2.7) on the left, we get

$$n \log\left(1 + \frac{r}{n}\right) + r \log\left(1 + \frac{n}{r}\right) \geq \frac{n \log n}{d} + O(1)$$

For all $x > 0$, we have $\log(1+x) \leq x$. So, $r \log\left(1+\frac{n}{r}\right) \leq n$,

$$\begin{aligned} \Rightarrow n \log\left(1+\frac{r}{n}\right) + n &\geq \frac{n \log n}{d} + O(1) \\ \log\left(1+\frac{r}{n}\right) &\geq \frac{\log n}{d} - 1 + O(1/n) \\ \frac{r}{n} &\geq n^{1/d}.e^{-1+O(1/n)} - 1 \end{aligned}$$

Assuming sufficiently large n such that $d \leq \log_3 n$ i.e $n^{1/d} \geq 3$,

$$\begin{aligned} \frac{r}{n} &= \Omega(n^{1/d}) \\ s &= \Omega(dn^{1+1/d}) \end{aligned}$$

This completes the proof of the theorem. □

When we substitute $d = 4$ in the above result, for depth-4 circuits, we get the lower bound of $\Omega(n^{5/4})$, where the input is the set of coefficients $a_1, a_2, \ldots, a_n$ i.e input length is $N = n \log n$. Hence, this provides an $\tilde{\Omega}(N^{5/4})$ lower bound for depth-4 arithmetic circuits. But it appears to us that some careful analysis of the above calculations can lead to an $\approx \tilde{\Omega}(N^{4/3})$ lower bound, which is the current best for general depth four circuits to our knowledge.

5.2 Ran Raz: Elusive Polynomial Functions

The next result we discuss is Ran Raz's *"Elusive functions and Lower Bounds for Arithmetic Circuits"* from 2010. It approaches circuit lower bounds by introducing the notion of *elusive polynomial mappings*. They show that providing examples of explicit polynomial functions, whose image is not contained in some predefined set, can imply significant arithmetic circuit lower bounds, for instance the superpolynomial lower bounds on the size of arithmetic circuits computing the permanent polynomial. The main lower bound result in [Raz10] is as stated below.

Theorem 3 ([Raz10]) *Let n be a prime number and $1 \leq d \leq (\log_2 n)/100$ be an integer. There exists an N-variate polynomial $\tilde{f} : \mathbb{F}^N \mapsto \mathbb{F}$ of degree $(5d+2)$ where $N = n.(5d+2)$, such that any depth-$\lfloor d/3 \rfloor$ arithmetic circuit computing $\tilde{f}$, over any field $\mathbb{F}$, is of size $\Omega(n^{1+1/(2d)})$.*

We present a closer analysis of Raz's proof for the special case of depth four circuits. We observe that the result in Theorem 3 for depth four circuits, has $d \geq 8$ and number of input variables

$N = n.(5d+2) = O(n)$. Therefore, the lower bound implied for depth four circuits, is $\Omega(N^{9/8})$. First, we restate a few definitions and terminology, borrowed from the original work by Raz, essential to their analysis.

Definition 5.4 (Polynomial Mappings) *Let $\mathbb{F}$ be any field, a polynomial mapping $f : \mathbb{F}^n \mapsto \mathbb{F}^m$ of degree r is the map defined by the m-tuple $f = (f_1, f_2, \ldots, f_m)$, where for all $i \in [m]$, $f_i \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ are n-variate polynomials of degree at most r.*

A mapping f is said to be a *multilinear* mapping, if all the polynomials $f_1, f_2, \ldots, f_n$ are multilinear polynomials (i.e. degree with respect to every variable is at most 1). A mapping f is said to be a *homogeneous* mapping of degree r, if all the polynomials $f_1, f_2, \ldots, f_n$ are homogeneous polynomials of total degree r.

We also define the set $\text{Image}(f) \subseteq \mathbb{F}^m$ as follows:

$$\text{Image}(f) := \{(f_1(a), f_2(a), \ldots, f_m(a)) \mid a \in \mathbb{F}^n\}.$$

Note that for $f_i \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ for all $i \in [m]$, $f = (f_1, f_2, \ldots, f_m)$ can also be thought as a mapping $f : K^n \mapsto K^m$ where $K \supset \mathbb{F}$ is a field extension, as $F[x_1, x_2, \ldots, x_n] \subseteq K[x_1, x_2, \ldots, x_n]$. Based on the above definition of polynomial mappings and their Image sets, we define the notion of *elusive* polynomial mappings.

Definition 5.5 (Elusive Mappings) *We say that a mapping $f : \mathbb{F}^n \mapsto \mathbb{F}^m$ "eludes" another mapping $\Psi : \mathbb{F}^s \mapsto \mathbb{F}^m$, if $\text{Image}(f) \nsubseteq \text{Image}(\Psi)$. If f eludes every mapping $\Psi : \mathbb{F}^s \mapsto \mathbb{F}^m$ of degree at most r, we say f is (s, r)-elusive.*

The principle idea behind Raz's result is to find explicit constructions of polynomial mappings that *elude* all polynomial mappings $\Psi : \mathbb{F}^s \mapsto \mathbb{F}^m$ of degree r. Here, the notion of explicit polynomial mappings is closely derived from explicit polynomial functions. A polynomial $f \in \mathbb{F}[x_1, x_2, \ldots, x_N]$ is said to be explicitly defined if it belongs to the class VNP. Recall the precise definition of VNP, where a polynomial $f \in \mathsf{VNP}$ if and only if there exists a polynomial $g \in \mathsf{VP}$ in $(n + w)$ variables, such that

$$f(x_1, x_2, \ldots, x_n) = \sum_{e_1, \ldots, e_w \in \{0,1\}} g(x_1, \ldots, x_n, e_1, \ldots, e_w)$$

where $w = \mathsf{poly}(n)$. Following the above definition closely, we define the notion of $\mathsf{poly}(n)$-definability for polynomial mappings $f : \mathbb{F}^n \mapsto \mathbb{F}^m$.

Definition 5.6 (poly(n)-Definable Mappings) *We say that a polynomial mapping is* poly(n)*-definable, if there exists* $w = \mathsf{poly}(n)$, $k = \lceil \log_2 m \rceil$, *and a polynomial* $g \in \mathsf{VP}$ *in* $(n + w + k)$ *variables, such that for every* $i \in [m]$,

$$f_i(x_1, x_2, \ldots, x_n) = \sum_{e_1,\ldots,e_w \in \{0,1\}} g(x_1, \ldots, x_n, e_1 \ldots, e_w, i_1, \ldots, i_k) \tag{5.2}$$

where $(i_k, \ldots, i_1)$ *is the binary representation of* $i - 1$.

Hereafter in the arguments, if we mention an *explicit* polynomial mapping, it would refer to a poly(n)-definable mapping. Also, similar to polynomial families in VNP, the mapping f is poly(n)-definable if there exists a deterministic polynomial time Turing machine that computes the coefficient of the monomial $x_1^{\gamma_1} \ldots x_n^{\gamma_n}$ in f_i, when $\gamma_1, \ldots, \gamma_n, i$ are given as inputs.

We define a structure called *circuit graphs*, associated with arithmetic circuits. For every arithmetic circuit C, the circuit graph G_C is the directed graph consisting of gates and edges from C, but excluding the labels (weights) on the edges. Thus, size of G_C is equal to the size of the circuit C, that is the number of edges in C. The depth of G_C is also similarly defined as for C. We associate the notion of *syntactic degree* with a circuit graph. The syntactic degree of a node in a circuit graph is inductively defined as follows:

- For a leaf node, the syntactic degree is 1 if it is labelled by an input variable, else it is 0.
- For a sum gate, it is the maximum of the syntactic degrees of its children.
- For a product gate, it is the sum of the syntactic degrees of its children.

The syntactic degree of the node corresponding to the output gate of the circuit, is called the syntactic degree of the circuit graph G_C. The notion of circuit graphs enables us to construct a single circuit to capture computation of all n-variate homogeneous polynomials of degree r, known as the *Universal Arithmetic Circuit*.

Definition 5.7 *A circuit* Φ *is called universal for all circuits with* n *inputs and* n *outputs, of size* s *and computing homogeneous polynomials of degree* r, *if the following holds: for every* n*-tuple of homogeneous* r *degree polynomials* $f_1(x_1, \ldots, x_n), \ldots, f_n(x_1, \ldots, x_n)$ *that can be computed simultaneously by a size* s *circuit, there exist another circuit* C *of size* s *and arbitrary depth, computing* $f_1, \ldots, f_n$ *such that the circuit graph* $G_C = G_\Phi$.

The existence of a universal circuit, for any $n, r, s \in \mathbb{N}$, was shown by [Raz10], which we use in the proof strategy for Theorem 3 below.

Proof Overview: Let $m = \binom{n+r-1}{r}$ be the number of monomials in n variables of degree exactly r. Then, every n-variate degree r polynomial $p \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ can be represented as a vector in $\mathbb{F}^m$, where every index corresponds to the coefficient of a r degree monomial in n variables. Hereafter, when we say a polynomial $p \in \mathbb{F}^m$, we refer to the homogeneous n-variate degree r polynomial represented by the vector p in $\mathbb{F}^m$ as explained above. Consider a universal arithmetic circuit U with inputs $x_1, \ldots, x_n$, for computing all possible homogeneous n-variate degree r polynomials, and let s be the size of the circuit U, i.e the number of edges in U. If we consider the labels of the edges in U as formal variables $\{y_1, \ldots, y_s\}$, then from the definition of universal arithmetic circuit it follows that the output of U is a polynomial $p \in \mathbb{F}^m$, where every entry of the vector is a polynomial in $\{y_1, \ldots, y_s\}$ variables. This computation is represented as a polynomial mapping $\Psi : \mathbb{F}^s \mapsto \mathbb{F}^m$. The proof of Theorem 3 involves three major steps:

- First, we prove the following: *Let a polynomial $g \in \mathbb{F}^m$ be computed by an arithmetic circuit of size s', then for every n, r, s', there exists a polynomial mapping $\Psi : \mathbb{F}^s \mapsto \mathbb{F}^m$ (where $s = \mathsf{poly}(s', n, r)$) such that $g \in \text{Image}(\Psi)$.* Raz proves the existence of the mapping Ψ, of degree $O(r)$ and that it can be constructed in time $\mathsf{poly}(s^r)$. In a way, Ψ captures the computation of all polynomials of 'low' complexity (complexity of size s'), and hence the task at hand is to find a set of polynomials in $\mathbb{F}^m$ not contained in $\text{Image}(\Psi)$, or in other words another polynomial mapping that *eludes* Ψ.

- Then the proof proceeds by showing an explicit ($\mathsf{poly}(n)$-definable) description of a polynomial mapping f that is (s, d)-elusive for appropriate s, d, and hence elusive of the mapping Ψ.

- Finally, Raz gives a 'hard' polynomial family $\tilde{f}$ derived from the elusive functions in mapping f, such that f being (s, d)-elusive implies that any arithmetic circuit computing $\tilde{f}$ must have large size.

Similar to [SS97], the lower bound proof by Raz focusses on a special form of arithmetic circuits, known as the *Normal Linear Form* of arithmetic circuits, that compute homogeneous linear polynomials (total degree one) also called linear forms.

Definition 5.8 (Normal Linear Form) *An arithmetic circuit is said to be in normal linear form if every intermediate gate is a sum gate, and all the leaf nodes are labelled by input variables (no field constants). Further, if we construct the circuit graph for an arithmetic circuit in normal linear form, the syntactic degree of every node is exactly one.*

It is also easy to observe the following result that implies equivalence of general arithmetic circuits computing linear forms, and arithmetic circuits in normal linear form.

Proposition 5.9 *Over any field $\mathbb{F}$, if C is an arithmetic circuit of size s and depth d computing n linear forms $f_1, f_2, \ldots f_n$, then there exists a circuit C' in normal linear form of size s and depth d, that also computes the polynomials $f_1, f_2, \ldots, f_n$.*

We describe the three steps that prove Theorem 3 below.

5.2.1 Description of Ψ

Let $\mathbb{F}$ be a field, n, r be integers such that $1 \leq r \leq n$ and n is a power of 2. Let $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ be the set of input variables, and M be the set of monomials in $\mathbf{x}$ of degree exactly r. Thus, $|M| = \binom{n+r-1}{r} = m'$. Let Γ be the set of all homogeneous r degree polynomials in $\mathbb{F}[\mathbf{x}]$, thus Γ is identified as the set $\mathbb{F}^{m'}$ as every polynomial in Γ is represented as a coefficient vector of dimension m'. Assume the following precedence among the $\mathbf{x}$-variables:$x_1 > x_2 > \ldots > x_n$. Naturally using this ordering among the variables, we can lexicographically order the monomials in M. This allows us to identify the set M with the set $[m']$ where $i \in [m']$ corresponds to the i-th monomial in the above lexicographical ordering. Further, consider the set Γ^n, of n-tuples $(g_1, g_2, \ldots, g_n)$ where every $g_i \in \Gamma$ for all $i \in [n]$. Every member of the set Γ^n can be represented as a vector of dimension $m = m'.n = n \cdot \binom{n+r-1}{r}$, concatenating the coefficient vectors of the n polynomials. Thus, every tuple in Γ^n can be identified by a vector in $\mathbb{F}^m$. This generates a homomorphism H from the set Γ^n to the set $\mathbb{F}^m$, which is used hereafter in the proof, to denote the n-tuples of homogeneous polynomials of degree r in $\mathbb{F}[\mathbf{x}]$. In other words, for every $(g_1, g_2, \ldots, g_n) \in \Gamma^n$, $H((g_1, g_2, \ldots, g_n)) \in \mathbb{F}^m$.

Consider the set $\mathcal{G}_{n,r}$ of circuit graphs that have n input gates labelled $x_1, \ldots, x_n$, and n output gates, all of syntactic degree r. Let $s \geq n$ be the size of every graph $G \in \mathcal{G}_{n,r}$. Let C be an arithmetic circuit over $\mathbb{F}$, with corresponding circuit graph $G_C \in \mathcal{G}_{n,r}$. Let the s edges of G_C be labelled by the variables $y_1, y_2, \ldots, y_s$. Then, C computes n homogeneous polynomials in $\mathbb{F}[\mathbf{x}]$, of degree exactly r, such that the coefficients of these polynomials are functions of $y_1, y_2, \ldots, y_s$. This gives us the polynomial mapping $\Psi_G : \mathbb{F}^s \mapsto \mathbb{F}^m$, where $\Psi_G(y_1, \ldots, y_s)$ is the m-tuple of polynomials $(\Psi_1, \ldots, \Psi_m)$ in $\mathbb{F}[y_1, \ldots, y_s]$ representing the coefficient of monomials in the n output polynomials of C. It is worth noting that the polynomials $\Psi_1, \ldots, \Psi_m$ are only dependent on the characteristic of the field $\mathbb{F}$, not on the field $\mathbb{F}$. (As the coefficients are only sums of products of $0, 1$-values that are contained in the minimal subfield of $\mathbb{F}$.) We make the following simple observation.

Proposition 5.10 *Let $G \in \mathcal{G}_{n,r}$ be a circuit graph. For every $g = (g_1, g_2, \ldots, g_n) \in \Gamma^n$, $H(g) \in \text{Image}(\Psi_G)$ if and only if $g_1, g_2, \ldots, g_n$ are computed by an arithmetic circuit C over $\mathbb{F}$*

such that $G_C = G$.

Proof: If $H(g) \in \text{Image}(\Psi_G)$, then by the definition of Ψ_G, there exists an arithmetic circuit C over $\mathbb{F}$, with $G_C = G$ that computes the polynomials $g_1, g_2, \ldots, g_n$ as outputs. For the other direction, let C be the arithmetic circuit over $\mathbb{F}$, computing the n polynomials $g_1, g_2, \ldots, g_n$ such that $G_C = G$. Since the circuit graph of G_C has s edges, the size of circuit C is s and suppose these s edges are labelled by $\alpha_1, \alpha_2, \ldots, \alpha_s \in \mathbb{F}$. Then, by definition of Ψ_G, $\Psi_G(\alpha_1, \ldots, \alpha_n) = H(g)$. □

Proposition 5.11 ([Raz10]) *Let $r = 1$, and hence $m = n^2$. If $G_C \in \mathcal{G}_{n,r}$ is the circuit graph of a circuit C in normal linear form, then $\Psi_G : \mathbb{F}^s \mapsto \mathbb{F}^m$ is a polynomial mapping of degree equal to the depth of the circuit C.*

Proof: Let C be the arithmetic circuit with circuit graph $G_C = G$ and its s edges be labelled by the variables $Y = \{y_1, \ldots, y_s\}$. Let v be a gate in C, and let $g_v \in \mathbb{F}[\mathbf{x}]$ be the intermediate polynomial computed at v. If v is a leaf, all the coefficients of g_v are independent of the variables $y_1, \ldots, y_s$. Using this as the base case, we use induction to prove that if v is at a distance d_v from the farthest leaf node, then every coefficient of the polynomial g_v is of degree at most d_v. The inductive proof is as follows: All children of the gate v are at distance $d_v - 1$ from their farthest leaves, and hence for every child u of v, the coefficients of the polynomial g_u are of degree at most $d_v - 1$ in the variables $y_1, \ldots, y_s$. Since v is an addition gate (normal linear form) and there is a single edge labelled by a Y-variable between v and any of its children, the degree of the coefficients of the polynomial g_v is at most d_v. When v is the output gate, the degree d is equal to the depth of the circuit C. □

It has been shown in [Raz10] that given n, r, s' ($s' \geq n$) as inputs, the universal arithmetic circuit on n inputs, computing n homogeneous polynomials of degree r, can be constructed in time $\mathsf{poly}(s, r)$ and has size $O(s'r^4)$. Further, [Raz10] also proves that we can conclude that given the graph G corresponding to the universal circuit, there exists a Turing machine that computes the polynomial mapping $\Psi_G : \mathbb{F}^s \mapsto \mathbb{F}^m$ of degree $O(r)$, where $s = O((s')^2 r^8)$. Further, from Propositions 5.10 and 5.11, this mapping Ψ_G captures computation of polynomials with 'low' complexity, as required.

5.2.2 Explicit Elusive Mapping f

Let n be a prime, $m = n^2$. Then, the set $[m]$ can be identified as $[n] \times [n]$ (in lexicographic ordering). Let $1 \leq d \leq (\log_2 n)/100$ be an integer. Then the set of input variables for the

polynomial mapping f is the set $X = \{x_{i,j}\}_{i\in[5d],j\in[n]}$. Hence, there are $n.(5d)$ input variables. For every $(a,b) \in [n] \times [n] = [m]$, define the polynomial

$$f_{(a,b)}(x_{1,1},\ldots,x_{5d,n}) := \prod_{i\in[5d]} x_{i,a+i.b}$$

where the sum $a + i.b$ is calculated modulo n. Thus, $f = (f_{(1,1)},\ldots,f_{(n,n)})$ is the m-tuple that defines the polynomial mapping $f : K^{n.(5d)} \mapsto K^m$ over any field K. The proof of the following lemma is omitted here and the interested reader may refer to [Raz10].

Lemma 5.12 ([Raz10]) *Let n be a prime. Let $m = n^2$ and $1 \leq d \leq (\log_2 n)/100$ be integers. Let K be a field of size at least m. Then, the polynomial mapping $f : K^{n.(5d)} \mapsto K^m$ is (s,d)-elusive, where $s = \lceil n^{1+1/(2d)} \rceil$.*

It is also easy to prove, as has been shown in [Raz10], that f is a $\mathsf{poly}(n)$-definable mapping as required.

5.2.3 The hard polynomial $\tilde{f}$

Let $\mathbf{z} = \{z_1,\ldots,z_n\}$ be an additional set of input variables. Then, define the following set of polynomials derived from the polynomial mapping f defined above. For every $i \in [n]$,

$$\tilde{f}_i(x_{1,1},\ldots,x_{5d,n},z_1,\ldots,z_n) := \sum_{j\in[n]} z_j \cdot f_{(i,j)}.$$

For every $a \in \mathbb{F}^{n.(5d)}$, we can substitute a for variables in X, and get $\tilde{f}_1|_a,\ldots,\tilde{f}_n|_a \in \mathbb{F}[\mathbf{z}]$. The following result has ben proved in [Raz10].

Proposition 5.13 *$\forall a \in \mathbb{F}^{n.(5d)}$, we have $(\tilde{f}_1|_a,\ldots,\tilde{f}_n|_a) \in \Gamma_n$ and $H((\tilde{f}_1|_a,\ldots,\tilde{f}_n|_a)) = f(a)$.*

In view of the above definitions, we prove the main result that connects the elusive mapping f to depth d arithmetic circuit size, and helps achieve the claimed lower bound.

Proposition 5.14 ([Raz10]) *Let m,n,d,s be integers such that $n,d \leq s$ and $m = n^2$. Let $f : \mathbb{F}^n \mapsto \mathbb{F}^m$ be a polynomial mapping. If there exists a field extension $K \supseteq \mathbb{F}$ such that f is (s,d)-elusive over K, then any depth-d arithmetic circuit over $\mathbb{F}$ computing the polynomials $\tilde{f}_1,\tilde{f}_2,\ldots,\tilde{f}_n : \mathbb{F}^{(5d+1).n} \mapsto \mathbb{F}$ must have size at least s.*

Proof: First, we prove the above result for $K = \mathbb{F}$. Let $\mathbb{F}(X)$ be the set of rational functions in the variables $\{x_{1,1},\ldots,x_{5d,n}\}$ over $\mathbb{F}$. By definition of $\tilde{f}_1,\tilde{f}_2,\ldots,\tilde{f}_n \in \mathbb{F}[X,\mathbf{z}]$ have $\mathbf{z}$-degree

1 and therefore total degree at most one more than the degree of f. Thus, the polynomials $\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_n$ are members of the polynomial ring $\mathbb{F}(X)[z_1, \ldots, z_n]$, in fact they are all linear polynomials in $\mathbf{z}$ variables with coefficients in $\mathbb{F}(X)$.

Suppose there exists an arithmetic circuit C, over $\mathbb{F}$, computing the polynomials $\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_n \in \mathbb{F}(X)[\mathbf{z}]$, of size s and depth d. The circuit C is computing linear polynomials in the variables $z_1, \ldots, z_n$, and hence we can assume it is in normal linear form (by Proposition 5.9). Further, as the proof of Proposition 5.9 involves elimination of the division operation, the labels of the edges in the normal linear form circuit, can be assumed to be in $\mathbb{F}[X]$ rather than $\mathbb{F}(X)$ (without loss of generality). Let $G = G_C$ be the circuit graph of the circuit C, in normal linear form. By Proposition 5.11, the mapping Ψ_G is a polynomial mapping of degree d. Since f is (s, d)-elusive, $\text{Image}(f) \nsubseteq \text{Image}(\Psi_G)$ and there exists a point $a \in \mathbb{F}^{n.(5d)}$ such that $f(a) \notin \text{Image}(\Psi_G)$. Substituting $x_{1,1} = a_{1,1}, \ldots, x_{5d,n} = a_{5d,n}$ in the circuit C gives us another circuit of size s and depth d (with circuit graph G), over $\mathbb{F}$, that computes the n polynomials $\tilde{f}_1|_a, \ldots, \tilde{f}_n|_a \in \mathbb{F}[\mathbf{z}]$. By Proposition 5.13, $H((\tilde{f}_1|_a, \ldots, \tilde{f}_n)) = f(a)$, and by the definition of Ψ_G (Proposition 5.10), $H((\tilde{f}_1|_a, \ldots, \tilde{f}_n)) \in \text{Image}(\Psi_G)$. This contradicts the assumption that $f(a) \notin \text{Image}(\Psi_G)$. Hence, the proposition is proved for $K = \mathbb{F}$.

For any general field extension $K \subseteq \mathbb{F}$, we assume that f is (s, d)-elusive over K. We proved above that any depth d circuit over K (where $K = \mathbb{F}$) for the polynomials $\tilde{(f)}_1, \ldots, \tilde{f}_n$ must have size at least s. But, an arithmetic circuit over $\mathbb{F}$ is also an arithmetic circuit over K. Hence, any depth d circuit over $\mathbb{F}$ for the polynomials $\tilde{(f)}_1, \ldots, \tilde{f}_n$ must have size at least s. □

5.2.4 Putting it together: Proof of Theorem 3

From Proposition 5.14 and Lemma 5.12 proved above, the following result is a direct implication.

Corollary 5.15 ([Raz10]) *Let n be a prime and $1 \leq d \leq (\log_2 n)/100$ be an integer. Any depth d arithmetic circuit, over any field $\mathbb{F}$ that computes the n polynomials $\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_n : \mathbb{F}^{n.(5d+1)} \mapsto \mathbb{F}$ (defined above), must have size at least $n^{1+1/(2d)}$.*

In order to define the single output polynomial for the result in Theorem 3, consider another additional set of variables $\{w_1, \ldots, w_n\}$. Now, define the polynomial $\tilde{f}$ of degree $(5d + 2)$ from the n polynomials $\tilde{f}_i$'s ($i \in [n]$) described above.

$$\tilde{f} = \sum_{k \in [n]} w_k \cdot \tilde{f}_k.$$

Observe that we get the polynomials $\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_n : \mathbb{F}^{n.(5d+1)} \mapsto \mathbb{F}$ as the set of first order partial derivatives of $\tilde{f}$ with respect to w_k variables ($k \in [n]$). Applying the result by [BS83] on the polynomial $\tilde{f}$, if $\tilde{f}$ is computed by a circuit of size s' and depth d, its n partial derivative polynomials are computed by an arithmetic circuit of size $5s'$ and depth $3d'$. Therefore, we get the result in Theorem 3. Substituting $d = 4$, the result is equal to a $\Omega(N^{9/8})$ lower bound for depth four arithmetic circuits.

5.3 Comparison with our result

As described in the earlier sections, we improve upon the implicit lower bounds for depth four circuits from the above two results. However, the two results are more closely related than they appear to be. On careful observation, [SS97] can be visualised as a special case of [Raz10]'s result, as the circuit provided for polynomial evaluation is essentially computing a polynomial mapping $\mathbb{F}^n \mapsto \mathbb{F}^{n^2}$ of degree $O(n)$, that is (s, r)-elusive for $s = n^{1+\Omega(1/r)}$. Hence, Raz generalizes the result in [SS97] for other kinds of polynomial functions. Further, the points for evaluation are not considered part of the input in [SS97], which is a significant variation from lower bound techniques used today.

On the other hand, we make use of recently successful techniques of partial derivative measures and Nisan-Wigderson Polynomials, to achieve an asymptotically stronger result of $\tilde{\Omega}(N^{3/2})$ as compared to $\Omega(N^{9/8})$ and $\tilde{\Omega}(N^{5/4})$ (rather $\approx \tilde{\Omega}(N^{4/3})$) by [Raz10] and [SS97] respectively. We provide an explicit polynomial family in VNP, and our circuit model is strictly $\Sigma\Pi\Sigma\Pi$ which is not restricted to linear circuits as in the above discussed results. Further, the hard polynomial used in [SS97] is a linear polynomial (total $\mathbf{x}$-degree is one), and the polynomial ised in [Raz10] has constant degree ($= 5d + 2$, for depth-$\lceil d/3 \rceil$ circuits), where as the degree is a function of N in our NW polynomial. This relation plays a major role in shaping our proof.

Bibliography

[Alo09] Noga Alon. Perturbed Identity Matrices Have High Rank: Proof and Applications. *Combinatorics, Probability & Computing*, 18(1-2):3–15, 2009. 6, 25

[AV08] Manindra Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *FOCS*, pages 67–75, 2008. 2, 3

[BC15] Suman K. Bera and Amit Chakrabarti. A depth-five lower bound for iterated matrix multiplication. In *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, pages 183–197, 2015. 3

[BLS16] Nikhil Balaji, Nutan Limaye, and Srikanth Srinivasan. An almost cubic lower bound for ΣΠΣ circuits computing a polynomial in VP. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:143, 2016. 3

[BS83] W. Baur and V. Strassen. The complexity of partial derivatives. *Theoretical Computer Science*, 22(3):317–330, 1983. ii, 4, 54

[CKW11] Xi Chen, Neeraj Kayal, and Avi Wigderson. Partial Derivatives in Arithmetic Complexity and Beyond. *Foundations and Trends in Theoretical Computer Science*, 6(1-2):1–138, 2011. 4

[Coo71] Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, STOC '71, pages 151–158, New York, NY, USA, 1971. ACM. 1

[FLMS14] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In *STOC*, pages 128–135, 2014. 2

[GKKS13a] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: A chasm at depth three. In *Foundations of Computer Science (FOCS)*, pages 578–587, 2013. 3

[GKKS13b] Ankit Gupta, Neeraj Kayal, Pritish Kamath, and Ramprasad Saptharishi. Approaching the chasm at depth four. In *Conference on Computational Complexity (CCC)*, 2013. 2

[Kal85] K. A. Kalorkoti. A lower bound for the formula size of rational functions. *SIAM Journal on Computing*, 14(3):678–687, 8 1985. ii, 3

[Kay12] Neeraj Kayal. An exponential lower bound for the sum of powers of bounded degree polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:81, 2012. 2, 3

[KLSS14] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An exponential lower bound for homogeneous depth four arithmetic formulas. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, 2014. ii, 2, 3, 6, 12, 22, 23, 24, 25

[Koi12] Pascal Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theor. Comput. Sci.*, 448:56–65, 2012. 2

[KS14] Mrinal Kumar and Shubhangi Saraf. On the power of homogeneous depth 4 arithmetic circuits. *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, 2014. 2

[KS15] Neeraj Kayal and Chandan Saha. Lower bounds for depth three arithmetic circuits with small bottom fanin. In *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, pages 158–208, 2015. ii, 3, 6, 22, 23, 24, 25, 27, 32

[KSS14] Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *STOC*, pages 146–153, 2014. 2, 11

[KST16] Neeraj Kayal, Chandan Saha, and Sébastien Tavenas. An almost cubic lower bound for depth three arithmetic circuits. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP)*, 2016. ii, 3, 17, 39, 40

[NW97] Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997. 3

[Pud94] P. Pudlák. Communication in bounded depth circuits. *Combinatorica*, 14(2):203–216, Jun 1994. 4

[Raz10] Ran Raz. Elusive functions and lower bounds for arithmetic circuits. *Theory of Computing*, 6(1):135–177, 2010. ii, iii, 4, 46, 48, 51, 52, 53, 54

[RS03] Ran Raz and Amir Shpilka. Lower bounds for matrix product in bounded depth circuits with arbitrary gates. *SIAM Journal on Computing*, 32(2):488–513, 2003. 4

[SS97] Victor Shoup and Roman Smolensky. Lower bounds for polynomial evaluation and interpolation problems. *Computational Complexity*, 6(4):301–311, 1997. ii, iii, 4, 41, 42, 44, 49, 54

[Str73] Volker Strassen. Vermeidung von divisionen. *Journal fr die reine und angewandte Mathematik*, 264:184–202, 1973. 4, 42

[SW99] Amir Shpilka and Avi Wigderson. Depth-3 arithmetic formulae over fields of characteristic zero. In *IEEE Conference on Computational Complexity*, 1999. ii, 3, 39

[SY10] Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5:207–388, March 2010. 3, 4

[Tav13] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. In *MFCS*, pages 813–824, 2013. 2

[Val79] L. G. Valiant. Completeness Classes in Algebra. In *STOC '79: Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 249–261, New York, NY, USA, 1979. ACM Press. 1, 2

[Val82] L. G. Valiant. Reducibility by Algebraic Projections. In *Logic and Algorithmic: an International Symposium held in honor of Ernst Specker*, volume 30 of *Monographies de l'Enseignement Mathémathique*, pages 365–380, 1982. 1, 2

[VSBR83] L.G. Valiant, S. Skyum, S. Berkowitz, and C. Rackoff. Fast parallel computation of polynomials using few processors. *SIAM Journal on Computing*, 12(4):641–644, 1983. 2

[Wig06] Avi Wigderson. P, np and mathematics - a computational complexity perspective. In *Proceedings of the ICM 06 (Madrid)*, 2006. 3

[Yau16] Morris Yau. Almost cubic bound for depth three circuits in VP. *Electronic Colloquium on Computational Complexity (ECCC)*, 2016. 3